

R0B11N - R Project

2023-04-26

Abstract

Machine learning and statistical regression models have become increasingly popular in predicting house prices. With the vast amount of data available in the real estate market, machine learning algorithms can extract patterns and relationships that may not be obvious to humans. By analyzing historical data on factors such as location, size, and amenities, regression models can accurately estimate the value of a property. These models can also incorporate external factors such as economic trends and changes in demographics to further improve their accuracy. As the real estate industry continues to evolve, the use of machine learning and statistical regression models is likely to become even more essential for accurate and efficient property valuation, as conducted in this dataset from Kaggle, with Kansas City's Housing Market.

Introduction

The contents of the data sets are primarily oriented on 21,598 Kansas City Homes as a singular CSV file packaged with 2,919 homes used as training models for the statistical regression model under train.csv and test.csv. Considered factors for analysis include: Date of Sale, Price, Bedrooms, Bathrooms, Living Room Size, Lot Size, Waterfront, Directional Views, Condition (linear), Grade (linear), Square Footage Above Ground, Square Footage at Base, Year Built, Year Renovated, Zipcode, Latitude, and Longitude.

By using R's data cleaning tools, and statistical visualization, the goal of the project is to construct a powerful and accurate KNN and Linear Regression model to analyze future potential housing trends, and the fluidity of the housing market.

Completing the task at hand would result in understanding the effects some factors which are considered in the regression model have on the future of home prices in Kansas City, where market analysis can have statistical backing for houses in a 2.4 million inhabitant metropolitan area.

Data

For our data, we used this dataset from Kaggle (<https://www.kaggle.com/competitions/house-prices-advanced-regression-techniques/data>) which includes the test, train, and sample submissions CSV files.

Loading Packages and Libraries

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.4
## v tibble  3.1.8      v dplyr  1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.2      v forcats 0.5.2
```

```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag() masks stats::lag()
```

```
library(corrplot)
```

```
## Warning: package 'corrplot' was built under R version 4.2.3
```

```
## corrplot 0.92 loaded
```

```
library(ggplot2)
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
##
## The following objects are masked from 'package:base':
##
##   date, intersect, setdiff, union
```

```
library(gridExtra)
```

```
## Warning: package 'gridExtra' was built under R version 4.2.3
```

```
##
## Attaching package: 'gridExtra'
##
## The following object is masked from 'package:dplyr':
##
##   combine
```

```
library(caTools)
```

```
## Warning: package 'caTools' was built under R version 4.2.3
```

```
library(GGally)
```

```
## Warning: package 'GGally' was built under R version 4.2.3
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg      ggplot2
```

Reading and Inspecting the Data

```
# Read The Training and Test Sets, Obtain All The Row and Column Data
test <- read_csv("test.csv")
```

```
## Rows: 1459 Columns: 80
## -- Column specification -----
## Delimiter: ","
## chr (43): MSZoning, Street, Alley, LotShape, LandContour, Utilities, LotConf...
## dbl (37): Id, MSSubClass, LotFrontage, LotArea, OverallQual, OverallCond, Ye...
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
train <- read_csv("train.csv")
```

```
## Rows: 1460 Columns: 81
## -- Column specification -----
## Delimiter: ","
## chr (43): MSZoning, Street, Alley, LotShape, LandContour, Utilities, LotConf...
## dbl (38): Id, MSSubClass, LotFrontage, LotArea, OverallQual, OverallCond, Ye...
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
sub_test <- read_csv("sample_submission.csv")
```

```
## Rows: 1459 Columns: 2
## -- Column specification -----
## Delimiter: ","
## dbl (2): Id, SalePrice
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
# Validate Structure of the Data Sets
```

```
head(train, 20)
```

```
## # A tibble: 20 x 81
##       Id MSSubClass MSZon~1 LotFr~2 LotArea Street Alley LotSh~3 LandC~4 Utili~5
##   <dbl>      <dbl> <chr>      <dbl>   <dbl> <chr>  <chr> <chr>   <chr>  <chr>
## 1     1         60 RL          65    8450 Pave  <NA>  Reg    Lvl    AllPub
## 2     2         20 RL          80    9600 Pave  <NA>  Reg    Lvl    AllPub
## 3     3         60 RL          68   11250 Pave  <NA>  IR1    Lvl    AllPub
## 4     4         70 RL          60    9550 Pave  <NA>  IR1    Lvl    AllPub
## 5     5         60 RL          84   14260 Pave  <NA>  IR1    Lvl    AllPub
## 6     6         50 RL          85   14115 Pave  <NA>  IR1    Lvl    AllPub
## 7     7         20 RL          75   10084 Pave  <NA>  Reg    Lvl    AllPub
## 8     8         60 RL          NA   10382 Pave  <NA>  IR1    Lvl    AllPub
## 9     9         50 RM          51    6120 Pave  <NA>  Reg    Lvl    AllPub
## 10    10        190 RL          50    7420 Pave  <NA>  Reg    Lvl    AllPub
## 11    11         20 RL          70   11200 Pave  <NA>  Reg    Lvl    AllPub
## 12    12         60 RL          85   11924 Pave  <NA>  IR1    Lvl    AllPub
## 13    13         20 RL          NA   12968 Pave  <NA>  IR2    Lvl    AllPub
## 14    14         20 RL          91   10652 Pave  <NA>  IR1    Lvl    AllPub
## 15    15         20 RL          NA   10920 Pave  <NA>  IR1    Lvl    AllPub
## 16    16         45 RM          51    6120 Pave  <NA>  Reg    Lvl    AllPub
```

```
## 17      17      20 RL      NA  11241 Pave <NA> IR1      Lvl      AllPub
## 18      18      90 RL      72  10791 Pave <NA> Reg      Lvl      AllPub
## 19      19      20 RL      66  13695 Pave <NA> Reg      Lvl      AllPub
## 20      20      20 RL      70   7560 Pave <NA> Reg      Lvl      AllPub
## # ... with 71 more variables: LotConfig <chr>, LandSlope <chr>,
## #   Neighborhood <chr>, Condition1 <chr>, Condition2 <chr>, BldgType <chr>,
## #   HouseStyle <chr>, OverallQual <dbl>, OverallCond <dbl>, YearBuilt <dbl>,
## #   YearRemodAdd <dbl>, RoofStyle <chr>, RoofMatl <chr>, Exterior1st <chr>,
## #   Exterior2nd <chr>, MasVnrType <chr>, MasVnrArea <dbl>, ExterQual <chr>,
## #   ExterCond <chr>, Foundation <chr>, BsmtQual <chr>, BsmtCond <chr>,
## #   BsmtExposure <chr>, BsmtFinType1 <chr>, BsmtFinSF1 <dbl>, ...
```

```
head(test, 20)
```

```
## # A tibble: 20 x 80
##       Id MSSubClass MSZon~1 LotFr~2 LotArea Street Alley LotSh~3 LandC~4 Utili~5
##       <dbl>      <dbl> <chr>      <dbl>   <dbl> <chr>  <chr> <chr> <chr> <chr>
## 1  1461         20 RH          80  11622 Pave <NA> Reg   Lvl   AllPub
## 2  1462         20 RL          81  14267 Pave <NA> IR1   Lvl   AllPub
## 3  1463         60 RL          74  13830 Pave <NA> IR1   Lvl   AllPub
## 4  1464         60 RL          78   9978 Pave <NA> IR1   Lvl   AllPub
## 5  1465        120 RL          43   5005 Pave <NA> IR1   HLS   AllPub
## 6  1466         60 RL          75  10000 Pave <NA> IR1   Lvl   AllPub
## 7  1467         20 RL          NA   7980 Pave <NA> IR1   Lvl   AllPub
## 8  1468         60 RL          63   8402 Pave <NA> IR1   Lvl   AllPub
## 9  1469         20 RL          85  10176 Pave <NA> Reg   Lvl   AllPub
## 10 1470         20 RL          70   8400 Pave <NA> Reg   Lvl   AllPub
## 11 1471        120 RH          26   5858 Pave <NA> IR1   Lvl   AllPub
## 12 1472        160 RM          21   1680 Pave <NA> Reg   Lvl   AllPub
## 13 1473        160 RM          21   1680 Pave <NA> Reg   Lvl   AllPub
## 14 1474        160 RL          24   2280 Pave <NA> Reg   Lvl   AllPub
## 15 1475        120 RL          24   2280 Pave <NA> Reg   Lvl   AllPub
## 16 1476         60 RL         102  12858 Pave <NA> IR1   Lvl   AllPub
## 17 1477         20 RL          94  12883 Pave <NA> IR1   Lvl   AllPub
## 18 1478         20 RL          90  11520 Pave <NA> Reg   Lvl   AllPub
## 19 1479         20 RL          79  14122 Pave <NA> IR1   Lvl   AllPub
## 20 1480         20 RL         110  14300 Pave <NA> Reg   HLS   AllPub
## # ... with 70 more variables: LotConfig <chr>, LandSlope <chr>,
## #   Neighborhood <chr>, Condition1 <chr>, Condition2 <chr>, BldgType <chr>,
## #   HouseStyle <chr>, OverallQual <dbl>, OverallCond <dbl>, YearBuilt <dbl>,
## #   YearRemodAdd <dbl>, RoofStyle <chr>, RoofMatl <chr>, Exterior1st <chr>,
## #   Exterior2nd <chr>, MasVnrType <chr>, MasVnrArea <dbl>, ExterQual <chr>,
## #   ExterCond <chr>, Foundation <chr>, BsmtQual <chr>, BsmtCond <chr>,
## #   BsmtExposure <chr>, BsmtFinType1 <chr>, BsmtFinSF1 <dbl>, ...
```

```
# Output Training and Test Sets as Data
str(test)
```

```
## spec_tbl_df [1,459 x 80] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ Id          : num [1:1459] 1461 1462 1463 1464 1465 ...
##  $ MSSubClass   : num [1:1459] 20 20 60 60 120 60 20 60 20 20 ...
##  $ MSZoning     : chr [1:1459] "RH" "RL" "RL" "RL" ...
##  $ LotFrontage  : num [1:1459] 80 81 74 78 43 75 NA 63 85 70 ...
```

```

## $ LotArea      : num [1:1459] 11622 14267 13830 9978 5005 ...
## $ Street       : chr [1:1459] "Pave" "Pave" "Pave" "Pave" ...
## $ Alley        : chr [1:1459] NA NA NA NA ...
## $ LotShape     : chr [1:1459] "Reg" "IR1" "IR1" "IR1" ...
## $ LandContour  : chr [1:1459] "Lvl" "Lvl" "Lvl" "Lvl" ...
## $ Utilities    : chr [1:1459] "AllPub" "AllPub" "AllPub" "AllPub" ...
## $ LotConfig    : chr [1:1459] "Inside" "Corner" "Inside" "Inside" ...
## $ LandSlope    : chr [1:1459] "Gtl" "Gtl" "Gtl" "Gtl" ...
## $ Neighborhood : chr [1:1459] "Names" "Names" "Gilbert" "Gilbert" ...
## $ Condition1   : chr [1:1459] "Feedr" "Norm" "Norm" "Norm" ...
## $ Condition2   : chr [1:1459] "Norm" "Norm" "Norm" "Norm" ...
## $ BldgType     : chr [1:1459] "1Fam" "1Fam" "1Fam" "1Fam" ...
## $ HouseStyle   : chr [1:1459] "1Story" "1Story" "2Story" "2Story" ...
## $ OverallQual  : num [1:1459] 5 6 5 6 8 6 6 6 7 4 ...
## $ OverallCond  : num [1:1459] 6 6 5 6 5 5 7 5 5 5 ...
## $ YearBuilt    : num [1:1459] 1961 1958 1997 1998 1992 ...
## $ YearRemodAdd : num [1:1459] 1961 1958 1998 1998 1992 ...
## $ RoofStyle    : chr [1:1459] "Gable" "Hip" "Gable" "Gable" ...
## $ RoofMatl     : chr [1:1459] "CompShg" "CompShg" "CompShg" "CompShg" ...
## $ Exterior1st  : chr [1:1459] "VinylSd" "Wd Sdng" "VinylSd" "VinylSd" ...
## $ Exterior2nd  : chr [1:1459] "VinylSd" "Wd Sdng" "VinylSd" "VinylSd" ...
## $ MasVnrType   : chr [1:1459] "None" "BrkFace" "None" "BrkFace" ...
## $ MasVnrArea   : num [1:1459] 0 108 0 20 0 0 0 0 0 0 ...
## $ ExterQual    : chr [1:1459] "TA" "TA" "TA" "TA" ...
## $ ExterCond    : chr [1:1459] "TA" "TA" "TA" "TA" ...
## $ Foundation   : chr [1:1459] "CBlock" "CBlock" "PConc" "PConc" ...
## $ BsmtQual     : chr [1:1459] "TA" "TA" "Gd" "TA" ...
## $ BsmtCond     : chr [1:1459] "TA" "TA" "TA" "TA" ...
## $ BsmtExposure : chr [1:1459] "No" "No" "No" "No" ...
## $ BsmtFinType1 : chr [1:1459] "Rec" "ALQ" "GLQ" "GLQ" ...
## $ BsmtFinSF1   : num [1:1459] 468 923 791 602 263 0 935 0 637 804 ...
## $ BsmtFinType2 : chr [1:1459] "LwQ" "Unf" "Unf" "Unf" ...
## $ BsmtFinSF2   : num [1:1459] 144 0 0 0 0 0 0 0 0 78 ...
## $ BsmtUnfSF    : num [1:1459] 270 406 137 324 1017 ...
## $ TotalBsmtSF  : num [1:1459] 882 1329 928 926 1280 ...
## $ Heating     : chr [1:1459] "GasA" "GasA" "GasA" "GasA" ...
## $ HeatingQC    : chr [1:1459] "TA" "TA" "Gd" "Ex" ...
## $ CentralAir   : chr [1:1459] "Y" "Y" "Y" "Y" ...
## $ Electrical   : chr [1:1459] "SBrkr" "SBrkr" "SBrkr" "SBrkr" ...
## $ 1stFlrSF     : num [1:1459] 896 1329 928 926 1280 ...
## $ 2ndFlrSF     : num [1:1459] 0 0 701 678 0 892 0 676 0 0 ...
## $ LowQualFinSF : num [1:1459] 0 0 0 0 0 0 0 0 0 0 ...
## $ GrLivArea    : num [1:1459] 896 1329 1629 1604 1280 ...
## $ BsmtFullBath : num [1:1459] 0 0 0 0 0 0 1 0 1 1 ...
## $ BsmtHalfBath : num [1:1459] 0 0 0 0 0 0 0 0 0 0 ...
## $ FullBath     : num [1:1459] 1 1 2 2 2 2 2 2 1 1 ...
## $ HalfBath     : num [1:1459] 0 1 1 1 0 1 0 1 1 0 ...
## $ BedroomAbvGr : num [1:1459] 2 3 3 3 2 3 3 3 2 2 ...
## $ KitchenAbvGr : num [1:1459] 1 1 1 1 1 1 1 1 1 1 ...
## $ KitchenQual  : chr [1:1459] "TA" "Gd" "TA" "Gd" ...
## $ TotRmsAbvGrd : num [1:1459] 5 6 6 7 5 7 6 7 5 4 ...
## $ Functional   : chr [1:1459] "Typ" "Typ" "Typ" "Typ" ...
## $ Fireplaces   : num [1:1459] 0 0 1 1 0 1 0 1 1 0 ...
## $ FireplaceQu  : chr [1:1459] NA NA "TA" "Gd" ...

```

```

## $ GarageType : chr [1:1459] "Attchd" "Attchd" "Attchd" "Attchd" ...
## $ GarageYrBlt : num [1:1459] 1961 1958 1997 1998 1992 ...
## $ GarageFinish : chr [1:1459] "Unf" "Unf" "Fin" "Fin" ...
## $ GarageCars : num [1:1459] 1 1 2 2 2 2 2 2 2 ...
## $ GarageArea : num [1:1459] 730 312 482 470 506 440 420 393 506 525 ...
## $ GarageQual : chr [1:1459] "TA" "TA" "TA" "TA" ...
## $ GarageCond : chr [1:1459] "TA" "TA" "TA" "TA" ...
## $ PavedDrive : chr [1:1459] "Y" "Y" "Y" "Y" ...
## $ WoodDeckSF : num [1:1459] 140 393 212 360 0 157 483 0 192 240 ...
## $ OpenPorchSF : num [1:1459] 0 36 34 36 82 84 21 75 0 0 ...
## $ EnclosedPorch: num [1:1459] 0 0 0 0 0 0 0 0 0 0 ...
## $ 3SsnPorch : num [1:1459] 0 0 0 0 0 0 0 0 0 0 ...
## $ ScreenPorch : num [1:1459] 120 0 0 0 144 0 0 0 0 0 ...
## $ PoolArea : num [1:1459] 0 0 0 0 0 0 0 0 0 0 ...
## $ PoolQC : chr [1:1459] NA NA NA NA ...
## $ Fence : chr [1:1459] "MnPrv" NA "MnPrv" NA ...
## $ MiscFeature : chr [1:1459] NA "Gar2" NA NA ...
## $ MiscVal : num [1:1459] 0 12500 0 0 0 0 500 0 0 0 ...
## $ MoSold : num [1:1459] 6 6 3 6 1 4 3 5 2 4 ...
## $ YrSold : num [1:1459] 2010 2010 2010 2010 2010 2010 2010 2010 2010 ...
## $ SaleType : chr [1:1459] "WD" "WD" "WD" "WD" ...
## $ SaleCondition: chr [1:1459] "Normal" "Normal" "Normal" "Normal" ...
## - attr(*, "spec")=
## .. cols(
## .. Id = col_double(),
## .. MSSubClass = col_double(),
## .. MSZoning = col_character(),
## .. LotFrontage = col_double(),
## .. LotArea = col_double(),
## .. Street = col_character(),
## .. Alley = col_character(),
## .. LotShape = col_character(),
## .. LandContour = col_character(),
## .. Utilities = col_character(),
## .. LotConfig = col_character(),
## .. LandSlope = col_character(),
## .. Neighborhood = col_character(),
## .. Condition1 = col_character(),
## .. Condition2 = col_character(),
## .. BldgType = col_character(),
## .. HouseStyle = col_character(),
## .. OverallQual = col_double(),
## .. OverallCond = col_double(),
## .. YearBuilt = col_double(),
## .. YearRemodAdd = col_double(),
## .. RoofStyle = col_character(),
## .. RoofMatl = col_character(),
## .. Exterior1st = col_character(),
## .. Exterior2nd = col_character(),
## .. MasVnrType = col_character(),
## .. MasVnrArea = col_double(),
## .. ExterQual = col_character(),
## .. ExterCond = col_character(),
## .. Foundation = col_character(),

```

```

## .. BsmtQual = col_character(),
## .. BsmtCond = col_character(),
## .. BsmtExposure = col_character(),
## .. BsmtFinType1 = col_character(),
## .. BsmtFinSF1 = col_double(),
## .. BsmtFinType2 = col_character(),
## .. BsmtFinSF2 = col_double(),
## .. BsmtUnfSF = col_double(),
## .. TotalBsmtSF = col_double(),
## .. Heating = col_character(),
## .. HeatingQC = col_character(),
## .. CentralAir = col_character(),
## .. Electrical = col_character(),
## .. '1stFlrSF' = col_double(),
## .. '2ndFlrSF' = col_double(),
## .. LowQualFinSF = col_double(),
## .. GrLivArea = col_double(),
## .. BsmtFullBath = col_double(),
## .. BsmtHalfBath = col_double(),
## .. FullBath = col_double(),
## .. HalfBath = col_double(),
## .. BedroomAbvGr = col_double(),
## .. KitchenAbvGr = col_double(),
## .. KitchenQual = col_character(),
## .. TotRmsAbvGrd = col_double(),
## .. Functional = col_character(),
## .. Fireplaces = col_double(),
## .. FireplaceQu = col_character(),
## .. GarageType = col_character(),
## .. GarageYrBlt = col_double(),
## .. GarageFinish = col_character(),
## .. GarageCars = col_double(),
## .. GarageArea = col_double(),
## .. GarageQual = col_character(),
## .. GarageCond = col_character(),
## .. PavedDrive = col_character(),
## .. WoodDeckSF = col_double(),
## .. OpenPorchSF = col_double(),
## .. EnclosedPorch = col_double(),
## .. '3SsnPorch' = col_double(),
## .. ScreenPorch = col_double(),
## .. PoolArea = col_double(),
## .. PoolQC = col_character(),
## .. Fence = col_character(),
## .. MiscFeature = col_character(),
## .. MiscVal = col_double(),
## .. MoSold = col_double(),
## .. YrSold = col_double(),
## .. SaleType = col_character(),
## .. SaleCondition = col_character()
## .. )
## - attr(*, "problems")=<externalptr>

```

```
str(train)
```

```
## spec_tbl_df [1,460 x 81] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ Id : num [1:1460] 1 2 3 4 5 6 7 8 9 10 ...
## $ MSSubClass : num [1:1460] 60 20 60 70 60 50 20 60 50 190 ...
## $ MSZoning : chr [1:1460] "RL" "RL" "RL" "RL" ...
## $ LotFrontage : num [1:1460] 65 80 68 60 84 85 75 NA 51 50 ...
## $ LotArea : num [1:1460] 8450 9600 11250 9550 14260 ...
## $ Street : chr [1:1460] "Pave" "Pave" "Pave" "Pave" ...
## $ Alley : chr [1:1460] NA NA NA NA ...
## $ LotShape : chr [1:1460] "Reg" "Reg" "IR1" "IR1" ...
## $ LandContour : chr [1:1460] "Lvl" "Lvl" "Lvl" "Lvl" ...
## $ Utilities : chr [1:1460] "AllPub" "AllPub" "AllPub" "AllPub" ...
## $ LotConfig : chr [1:1460] "Inside" "FR2" "Inside" "Corner" ...
## $ LandSlope : chr [1:1460] "Gtl" "Gtl" "Gtl" "Gtl" ...
## $ Neighborhood : chr [1:1460] "CollgCr" "Veenker" "CollgCr" "Crawfor" ...
## $ Condition1 : chr [1:1460] "Norm" "Feedr" "Norm" "Norm" ...
## $ Condition2 : chr [1:1460] "Norm" "Norm" "Norm" "Norm" ...
## $ BldgType : chr [1:1460] "1Fam" "1Fam" "1Fam" "1Fam" ...
## $ HouseStyle : chr [1:1460] "2Story" "1Story" "2Story" "2Story" ...
## $ OverallQual : num [1:1460] 7 6 7 7 8 5 8 7 7 5 ...
## $ OverallCond : num [1:1460] 5 8 5 5 5 5 5 6 5 6 ...
## $ YearBuilt : num [1:1460] 2003 1976 2001 1915 2000 ...
## $ YearRemodAdd : num [1:1460] 2003 1976 2002 1970 2000 ...
## $ RoofStyle : chr [1:1460] "Gable" "Gable" "Gable" "Gable" ...
## $ RoofMatl : chr [1:1460] "CompShg" "CompShg" "CompShg" "CompShg" ...
## $ Exterior1st : chr [1:1460] "VinylSd" "MetalSd" "VinylSd" "Wd Sdng" ...
## $ Exterior2nd : chr [1:1460] "VinylSd" "MetalSd" "VinylSd" "Wd Shng" ...
## $ MasVnrType : chr [1:1460] "BrkFace" "None" "BrkFace" "None" ...
## $ MasVnrArea : num [1:1460] 196 0 162 0 350 0 186 240 0 0 ...
## $ ExterQual : chr [1:1460] "Gd" "TA" "Gd" "TA" ...
## $ ExterCond : chr [1:1460] "TA" "TA" "TA" "TA" ...
## $ Foundation : chr [1:1460] "PConc" "CBlock" "PConc" "BrkTil" ...
## $ BsmtQual : chr [1:1460] "Gd" "Gd" "Gd" "TA" ...
## $ BsmtCond : chr [1:1460] "TA" "TA" "TA" "Gd" ...
## $ BsmtExposure : chr [1:1460] "No" "Gd" "Mn" "No" ...
## $ BsmtFinType1 : chr [1:1460] "GLQ" "ALQ" "GLQ" "ALQ" ...
## $ BsmtFinSF1 : num [1:1460] 706 978 486 216 655 ...
## $ BsmtFinType2 : chr [1:1460] "Unf" "Unf" "Unf" "Unf" ...
## $ BsmtFinSF2 : num [1:1460] 0 0 0 0 0 0 0 32 0 0 ...
## $ BsmtUnfSF : num [1:1460] 150 284 434 540 490 64 317 216 952 140 ...
## $ TotalBsmtSF : num [1:1460] 856 1262 920 756 1145 ...
## $ Heating : chr [1:1460] "GasA" "GasA" "GasA" "GasA" ...
## $ HeatingQC : chr [1:1460] "Ex" "Ex" "Ex" "Gd" ...
## $ CentralAir : chr [1:1460] "Y" "Y" "Y" "Y" ...
## $ Electrical : chr [1:1460] "SBrkr" "SBrkr" "SBrkr" "SBrkr" ...
## $ 1stFlrSF : num [1:1460] 856 1262 920 961 1145 ...
## $ 2ndFlrSF : num [1:1460] 854 0 866 756 1053 ...
## $ LowQualFinSF : num [1:1460] 0 0 0 0 0 0 0 0 0 0 ...
## $ GrLivArea : num [1:1460] 1710 1262 1786 1717 2198 ...
## $ BsmtFullBath : num [1:1460] 1 0 1 1 1 1 1 1 0 1 ...
## $ BsmtHalfBath : num [1:1460] 0 1 0 0 0 0 0 0 0 0 ...
## $ FullBath : num [1:1460] 2 2 2 1 2 1 2 2 2 1 ...
```



```

## $ HalfBath      : num [1:1460] 1 0 1 0 1 1 0 1 0 0 ...
## $ BedroomAbvGr : num [1:1460] 3 3 3 3 4 1 3 3 2 2 ...
## $ KitchenAbvGr : num [1:1460] 1 1 1 1 1 1 1 1 2 2 ...
## $ KitchenQual   : chr [1:1460] "Gd" "TA" "Gd" "Gd" ...
## $ TotRmsAbvGrd  : num [1:1460] 8 6 6 7 9 5 7 7 8 5 ...
## $ Functional    : chr [1:1460] "Typ" "Typ" "Typ" "Typ" ...
## $ Fireplaces    : num [1:1460] 0 1 1 1 1 0 1 2 2 2 ...
## $ FireplaceQu   : chr [1:1460] NA "TA" "TA" "Gd" ...
## $ GarageType    : chr [1:1460] "Attchd" "Attchd" "Attchd" "Detchd" ...
## $ GarageYrBlt   : num [1:1460] 2003 1976 2001 1998 2000 ...
## $ GarageFinish  : chr [1:1460] "RFn" "RFn" "RFn" "Unf" ...
## $ GarageCars    : num [1:1460] 2 2 2 3 3 2 2 2 2 1 ...
## $ GarageArea    : num [1:1460] 548 460 608 642 836 480 636 484 468 205 ...
## $ GarageQual    : chr [1:1460] "TA" "TA" "TA" "TA" ...
## $ GarageCond    : chr [1:1460] "TA" "TA" "TA" "TA" ...
## $ PavedDrive    : chr [1:1460] "Y" "Y" "Y" "Y" ...
## $ WoodDeckSF    : num [1:1460] 0 298 0 0 192 40 255 235 90 0 ...
## $ OpenPorchSF   : num [1:1460] 61 0 42 35 84 30 57 204 0 4 ...
## $ EnclosedPorch : num [1:1460] 0 0 0 272 0 0 0 228 205 0 ...
## $ 3SsnPorch     : num [1:1460] 0 0 0 0 0 320 0 0 0 0 ...
## $ ScreenPorch   : num [1:1460] 0 0 0 0 0 0 0 0 0 0 ...
## $ PoolArea      : num [1:1460] 0 0 0 0 0 0 0 0 0 0 ...
## $ PoolQC        : chr [1:1460] NA NA NA NA ...
## $ Fence         : chr [1:1460] NA NA NA NA ...
## $ MiscFeature   : chr [1:1460] NA NA NA NA ...
## $ MiscVal       : num [1:1460] 0 0 0 0 0 700 0 350 0 0 ...
## $ MoSold        : num [1:1460] 2 5 9 2 12 10 8 11 4 1 ...
## $ YrSold        : num [1:1460] 2008 2007 2008 2006 2008 ...
## $ SaleType      : chr [1:1460] "WD" "WD" "WD" "WD" ...
## $ SaleCondition : chr [1:1460] "Normal" "Normal" "Normal" "Abnorml" ...
## $ SalePrice     : num [1:1460] 208500 181500 223500 140000 250000 ...
## - attr(*, "spec")=
## .. cols(
## ..   Id = col_double(),
## ..   MSSubClass = col_double(),
## ..   MSZoning = col_character(),
## ..   LotFrontage = col_double(),
## ..   LotArea = col_double(),
## ..   Street = col_character(),
## ..   Alley = col_character(),
## ..   LotShape = col_character(),
## ..   LandContour = col_character(),
## ..   Utilities = col_character(),
## ..   LotConfig = col_character(),
## ..   LandSlope = col_character(),
## ..   Neighborhood = col_character(),
## ..   Condition1 = col_character(),
## ..   Condition2 = col_character(),
## ..   BldgType = col_character(),
## ..   HouseStyle = col_character(),
## ..   OverallQual = col_double(),
## ..   OverallCond = col_double(),
## ..   YearBuilt = col_double(),
## ..   YearRemodAdd = col_double(),

```

```

## .. RoofStyle = col_character(),
## .. RoofMatl = col_character(),
## .. Exterior1st = col_character(),
## .. Exterior2nd = col_character(),
## .. MasVnrType = col_character(),
## .. MasVnrArea = col_double(),
## .. ExterQual = col_character(),
## .. ExterCond = col_character(),
## .. Foundation = col_character(),
## .. BsmtQual = col_character(),
## .. BsmtCond = col_character(),
## .. BsmtExposure = col_character(),
## .. BsmtFinType1 = col_character(),
## .. BsmtFinSF1 = col_double(),
## .. BsmtFinType2 = col_character(),
## .. BsmtFinSF2 = col_double(),
## .. BsmtUnfSF = col_double(),
## .. TotalBsmtSF = col_double(),
## .. Heating = col_character(),
## .. HeatingQC = col_character(),
## .. CentralAir = col_character(),
## .. Electrical = col_character(),
## .. '1stFlrSF' = col_double(),
## .. '2ndFlrSF' = col_double(),
## .. LowQualFinSF = col_double(),
## .. GrLivArea = col_double(),
## .. BsmtFullBath = col_double(),
## .. BsmtHalfBath = col_double(),
## .. FullBath = col_double(),
## .. HalfBath = col_double(),
## .. BedroomAbvGr = col_double(),
## .. KitchenAbvGr = col_double(),
## .. KitchenQual = col_character(),
## .. TotRmsAbvGrd = col_double(),
## .. Functional = col_character(),
## .. Fireplaces = col_double(),
## .. FireplaceQu = col_character(),
## .. GarageType = col_character(),
## .. GarageYrBlt = col_double(),
## .. GarageFinish = col_character(),
## .. GarageCars = col_double(),
## .. GarageArea = col_double(),
## .. GarageQual = col_character(),
## .. GarageCond = col_character(),
## .. PavedDrive = col_character(),
## .. WoodDeckSF = col_double(),
## .. OpenPorchSF = col_double(),
## .. EnclosedPorch = col_double(),
## .. '3SsnPorch' = col_double(),
## .. ScreenPorch = col_double(),
## .. PoolArea = col_double(),
## .. PoolQC = col_character(),
## .. Fence = col_character(),
## .. MiscFeature = col_character(),

```

```
## .. MiscVal = col_double(),
## .. MoSold = col_double(),
## .. YrSold = col_double(),
## .. SaleType = col_character(),
## .. SaleCondition = col_character(),
## .. SalePrice = col_double()
## .. )
## - attr(*, "problems")=<externalptr>
```

```
summary(test)
```

```
##      Id      MSSubClass      MSZoning      LotFrontage
## Min.   :1461  Min.    : 20.00  Length:1459  Min.    : 21.00
## 1st Qu.:1826  1st Qu.: 20.00  Class :character  1st Qu.: 58.00
## Median :2190  Median : 50.00  Mode  :character  Median : 67.00
## Mean   :2190  Mean   : 57.38                Mean   : 68.58
## 3rd Qu.:2554  3rd Qu.: 70.00                3rd Qu.: 80.00
## Max.   :2919  Max.    :190.00                Max.    :200.00
##                                     NA's    :227
##      LotArea      Street      Alley      LotShape
## Min.   : 1470  Length:1459  Length:1459  Length:1459
## 1st Qu.: 7391  Class :character  Class :character  Class :character
## Median : 9399  Mode  :character  Mode  :character  Mode  :character
## Mean    : 9819
## 3rd Qu.:11518
## Max.    :56600
##
##      LandContour      Utilities      LotConfig      LandSlope
## Length:1459      Length:1459      Length:1459      Length:1459
## Class :character  Class :character  Class :character  Class :character
## Mode  :character  Mode  :character  Mode  :character  Mode  :character
##
##
##
##      Neighborhood      Condition1      Condition2      BldgType
## Length:1459      Length:1459      Length:1459      Length:1459
## Class :character  Class :character  Class :character  Class :character
## Mode  :character  Mode  :character  Mode  :character  Mode  :character
##
##
##
##      HouseStyle      OverallQual      OverallCond      YearBuilt
## Length:1459      Min.    : 1.000  Min.    :1.000  Min.    :1879
## Class :character  1st Qu.: 5.000  1st Qu.:5.000  1st Qu.:1953
## Mode  :character  Median : 6.000  Median :5.000  Median :1973
##                                     Mean   : 6.079  Mean   :5.554  Mean   :1971
##                                     3rd Qu.: 7.000  3rd Qu.:6.000  3rd Qu.:2001
##                                     Max.    :10.000  Max.    :9.000  Max.    :2010
##
##      YearRemodAdd      RoofStyle      RoofMatl      Exterior1st
## Min.    :1950  Length:1459  Length:1459  Length:1459
## 1st Qu.:1963  Class :character  Class :character  Class :character
```

```

## Median :1992   Mode  :character   Mode  :character   Mode  :character
## Mean    :1984
## 3rd Qu.:2004
## Max.    :2010
##
## Exterior2nd      MasVnrType      MasVnrArea      ExterQual
## Length:1459      Length:1459      Min.   :    0.0   Length:1459
## Class :character  Class :character  1st Qu.:    0.0   Class :character
## Mode  :character  Mode  :character  Median :    0.0   Mode  :character
##                                     Mean  : 100.7
##                                     3rd Qu.: 164.0
##                                     Max.   :1290.0
##                                     NA's   :15
## ExterCond      Foundation      BsmtQual      BsmtCond
## Length:1459      Length:1459      Length:1459      Length:1459
## Class :character  Class :character  Class :character  Class :character
## Mode  :character  Mode  :character  Mode  :character  Mode  :character
##
##
##
## BsmtExposure      BsmtFinType1      BsmtFinSF1      BsmtFinType2
## Length:1459      Length:1459      Min.   :    0.0   Length:1459
## Class :character  Class :character  1st Qu.:    0.0   Class :character
## Mode  :character  Mode  :character  Median : 350.5   Mode  :character
##                                     Mean  : 439.2
##                                     3rd Qu.: 753.5
##                                     Max.   :4010.0
##                                     NA's   :1
## BsmtFinSF2      BsmtUnfSF      TotalBsmtSF      Heating
## Min.   :    0.00   Min.   :    0.0   Min.   :    0   Length:1459
## 1st Qu.:    0.00   1st Qu.: 219.2   1st Qu.: 784   Class :character
## Median :    0.00   Median : 460.0   Median : 988   Mode  :character
## Mean    :   52.62   Mean    : 554.3   Mean    :1046
## 3rd Qu.:    0.00   3rd Qu.: 797.8   3rd Qu.:1305
## Max.    :1526.00   Max.    :2140.0   Max.    :5095
## NA's    :1        NA's    :1        NA's    :1
## HeatingQC      CentralAir      Electrical      1stFlrSF
## Length:1459      Length:1459      Length:1459      Min.   : 407.0
## Class :character  Class :character  Class :character  1st Qu.: 873.5
## Mode  :character  Mode  :character  Mode  :character  Median :1079.0
##                                     Mean    :1156.5
##                                     3rd Qu.:1382.5
##                                     Max.    :5095.0
##
## 2ndFlrSF      LowQualFinSF      GrLivArea      BsmtFullBath
## Min.   :    0   Min.   :    0.000   Min.   : 407   Min.   :0.0000
## 1st Qu.:    0   1st Qu.:    0.000   1st Qu.:1118   1st Qu.:0.0000
## Median :    0   Median :    0.000   Median :1432   Median :0.0000
## Mean    : 326   Mean    :    3.543   Mean    :1486   Mean    :0.4345
## 3rd Qu.: 676   3rd Qu.:    0.000   3rd Qu.:1721   3rd Qu.:1.0000
## Max.    :1862   Max.    :1064.000   Max.    :5095   Max.    :3.0000
##                                     NA's    :2
## BsmtHalfBath      FullBath      HalfBath      BedroomAbvGr

```

```

## Min. :0.0000 Min. :0.000 Min. :0.0000 Min. :0.000
## 1st Qu.:0.0000 1st Qu.:1.000 1st Qu.:0.0000 1st Qu.:2.000
## Median :0.0000 Median :2.000 Median :0.0000 Median :3.000
## Mean :0.0652 Mean :1.571 Mean :0.3777 Mean :2.854
## 3rd Qu.:0.0000 3rd Qu.:2.000 3rd Qu.:1.0000 3rd Qu.:3.000
## Max. :2.0000 Max. :4.000 Max. :2.0000 Max. :6.000
## NA's :2
## KitchenAbvGr KitchenQual TotRmsAbvGrd Functional
## Min. :0.000 Length:1459 Min. : 3.000 Length:1459
## 1st Qu.:1.000 Class :character 1st Qu.: 5.000 Class :character
## Median :1.000 Mode :character Median : 6.000 Mode :character
## Mean :1.042 Mean : 6.385
## 3rd Qu.:1.000 3rd Qu.: 7.000
## Max. :2.000 Max. :15.000
##
## Fireplaces FireplaceQu GarageType GarageYrBlt
## Min. :0.0000 Length:1459 Length:1459 Min. :1895
## 1st Qu.:0.0000 Class :character Class :character 1st Qu.:1959
## Median :0.0000 Mode :character Mode :character Median :1979
## Mean :0.5812 Mean :1978
## 3rd Qu.:1.0000 3rd Qu.:2002
## Max. :4.0000 Max. :2207
## NA's :78
## GarageFinish GarageCars GarageArea GarageQual
## Length:1459 Min. :0.000 Min. : 0.0 Length:1459
## Class :character 1st Qu.:1.000 1st Qu.: 318.0 Class :character
## Mode :character Median :2.000 Median : 480.0 Mode :character
## Mean :1.766 Mean : 472.8
## 3rd Qu.:2.000 3rd Qu.: 576.0
## Max. :5.000 Max. :1488.0
## NA's :1 NA's :1
## GarageCond PavedDrive WoodDeckSF OpenPorchSF
## Length:1459 Length:1459 Min. : 0.00 Min. : 0.00
## Class :character Class :character 1st Qu.: 0.00 1st Qu.: 0.00
## Mode :character Mode :character Median : 0.00 Median : 28.00
## Mean : 93.17 Mean : 48.31
## 3rd Qu.:168.00 3rd Qu.: 72.00
## Max. :1424.00 Max. :742.00
##
## EnclosedPorch 3SsnPorch ScreenPorch PoolArea
## Min. : 0.00 Min. : 0.000 Min. : 0.00 Min. : 0.000
## 1st Qu.: 0.00 1st Qu.: 0.000 1st Qu.: 0.00 1st Qu.: 0.000
## Median : 0.00 Median : 0.000 Median : 0.00 Median : 0.000
## Mean : 24.24 Mean : 1.794 Mean : 17.06 Mean : 1.744
## 3rd Qu.: 0.00 3rd Qu.: 0.000 3rd Qu.: 0.00 3rd Qu.: 0.000
## Max. :1012.00 Max. :360.000 Max. :576.00 Max. :800.000
##
## PoolQC Fence MiscFeature MiscVal
## Length:1459 Length:1459 Length:1459 Min. : 0.00
## Class :character Class :character Class :character 1st Qu.: 0.00
## Mode :character Mode :character Mode :character Median : 0.00
## Mean : 58.17
## 3rd Qu.: 0.00
## Max. :17000.00

```

```
##
##      MoSold      YrSold      SaleType      SaleCondition
## Min.   : 1.000   Min.   :2006   Length:1459   Length:1459
## 1st Qu.: 4.000   1st Qu.:2007   Class :character   Class :character
## Median : 6.000   Median :2008   Mode  :character   Mode  :character
## Mean   : 6.104   Mean    :2008
## 3rd Qu.: 8.000   3rd Qu.:2009
## Max.   :12.000   Max.    :2010
##
```

```
summary(train)
```

```
##      Id      MSSubClass      MSZoning      LotFrontage
## Min.   : 1.0   Min.   : 20.0   Length:1460   Min.   : 21.00
## 1st Qu.: 365.8 1st Qu.: 20.0   Class :character   1st Qu.: 59.00
## Median : 730.5 Median : 50.0   Mode  :character   Median : 69.00
## Mean   : 730.5 Mean   : 56.9
## 3rd Qu.:1095.2 3rd Qu.: 70.0
## Max.   :1460.0 Max.   :190.0
##              NA's   :259
##      LotArea      Street      Alley      LotShape
## Min.   : 1300   Length:1460   Length:1460   Length:1460
## 1st Qu.: 7554   Class :character   Class :character   Class :character
## Median : 9478   Mode  :character   Mode  :character   Mode  :character
## Mean   : 10517
## 3rd Qu.: 11602
## Max.   :215245
##
##      LandContour      Utilities      LotConfig      LandSlope
## Length:1460      Length:1460      Length:1460      Length:1460
## Class :character   Class :character   Class :character   Class :character
## Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##      Neighborhood      Condition1      Condition2      BldgType
## Length:1460      Length:1460      Length:1460      Length:1460
## Class :character   Class :character   Class :character   Class :character
## Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##      HouseStyle      OverallQual      OverallCond      YearBuilt
## Length:1460      Min.   : 1.000   Min.   :1.000   Min.   :1872
## Class :character   1st Qu.: 5.000   1st Qu.:5.000   1st Qu.:1954
## Mode  :character   Median : 6.000   Median :5.000   Median :1973
##              Mean   : 6.099   Mean   :5.575   Mean   :1971
##              3rd Qu.: 7.000   3rd Qu.:6.000   3rd Qu.:2000
##              Max.   :10.000   Max.   :9.000   Max.   :2010
##
##      YearRemodAdd      RoofStyle      RoofMatl      Exterior1st
## Min.   :1950   Length:1460      Length:1460      Length:1460
```

```

## 1st Qu.:1967   Class :character   Class :character   Class :character
## Median :1994   Mode  :character   Mode  :character   Mode  :character
## Mean    :1985
## 3rd Qu.:2004
## Max.    :2010
##
## Exterior2nd      MasVnrType      MasVnrArea      ExterQual
## Length:1460      Length:1460      Min.   :  0.0      Length:1460
## Class :character   Class :character  1st Qu.:  0.0      Class :character
## Mode  :character   Mode  :character  Median :  0.0      Mode  :character
##                                     Mean   : 103.7
##                                     3rd Qu.: 166.0
##                                     Max.   :1600.0
##                                     NA's   :8
## ExterCond      Foundation      BsmtQual      BsmtCond
## Length:1460      Length:1460      Length:1460      Length:1460
## Class :character   Class :character   Class :character   Class :character
## Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
## BsmtExposure      BsmtFinType1      BsmtFinSF1      BsmtFinType2
## Length:1460      Length:1460      Min.   :  0.0      Length:1460
## Class :character   Class :character   1st Qu.:  0.0      Class :character
## Mode  :character   Mode  :character   Median : 383.5      Mode  :character
##                                     Mean   : 443.6
##                                     3rd Qu.: 712.2
##                                     Max.   :5644.0
##
## BsmtFinSF2      BsmtUnfSF      TotalBsmtSF      Heating
## Min.   :  0.00      Min.   :  0.0      Min.   :  0.0      Length:1460
## 1st Qu.:  0.00      1st Qu.: 223.0      1st Qu.: 795.8      Class :character
## Median :  0.00      Median : 477.5      Median : 991.5      Mode  :character
## Mean   :  46.55      Mean   : 567.2      Mean   :1057.4
## 3rd Qu.:  0.00      3rd Qu.: 808.0      3rd Qu.:1298.2
## Max.   :1474.00      Max.   :2336.0      Max.   :6110.0
##
## HeatingQC      CentralAir      Electrical      1stFlrSF
## Length:1460      Length:1460      Length:1460      Min.   : 334
## Class :character   Class :character   Class :character   1st Qu.: 882
## Mode  :character   Mode  :character   Mode  :character   Median :1087
##                                     Mean   :1163
##                                     3rd Qu.:1391
##                                     Max.   :4692
##
## 2ndFlrSF      LowQualFinSF      GrLivArea      BsmtFullBath
## Min.   :  0      Min.   :  0.000      Min.   : 334      Min.   :0.0000
## 1st Qu.:  0      1st Qu.:  0.000      1st Qu.:1130      1st Qu.:0.0000
## Median :  0      Median :  0.000      Median :1464      Median :0.0000
## Mean   : 347      Mean   :  5.845      Mean   :1515      Mean   :0.4253
## 3rd Qu.: 728      3rd Qu.:  0.000      3rd Qu.:1777      3rd Qu.:1.0000
## Max.   :2065      Max.   :572.000      Max.   :5642      Max.   :3.0000
##

```

```

##      BsmtHalfBath      FullBath      HalfBath      BedroomAbvGr
## Min.      :0.00000  Min.      :0.000  Min.      :0.0000  Min.      :0.000
## 1st Qu.:0.00000  1st Qu.:1.000  1st Qu.:0.0000  1st Qu.:2.000
## Median :0.00000  Median :2.000  Median :0.0000  Median :3.000
## Mean   :0.05753  Mean   :1.565  Mean   :0.3829  Mean   :2.866
## 3rd Qu.:0.00000  3rd Qu.:2.000  3rd Qu.:1.0000  3rd Qu.:3.000
## Max.   :2.00000  Max.   :3.000  Max.   :2.0000  Max.   :8.000
##
##      KitchenAbvGr  KitchenQual      TotRmsAbvGrd      Functional
## Min.      :0.000  Length:1460  Min.      : 2.000  Length:1460
## 1st Qu.:1.000  Class :character  1st Qu.: 5.000  Class :character
## Median :1.000  Mode  :character  Median : 6.000  Mode  :character
## Mean   :1.047
## 3rd Qu.:1.000
## Max.   :3.000
##
##      Fireplaces  FireplaceQu      GarageType      GarageYrBlt
## Min.      :0.000  Length:1460  Length:1460  Min.      :1900
## 1st Qu.:0.000  Class :character  Class :character  1st Qu.:1961
## Median :1.000  Mode  :character  Mode  :character  Median :1980
## Mean   :0.613
## 3rd Qu.:1.000
## Max.   :3.000
##
##      GarageFinish      GarageCars      GarageArea      GarageQual
## Length:1460  Min.      :0.000  Min.      : 0.0  Length:1460
## Class :character  1st Qu.:1.000  1st Qu.: 334.5  Class :character
## Mode  :character  Median :2.000  Median : 480.0  Mode  :character
##
##      Mean   :1.767  Mean   : 473.0
##      3rd Qu.:2.000  3rd Qu.: 576.0
##      Max.   :4.000  Max.   :1418.0
##
##      GarageCond      PavedDrive      WoodDeckSF      OpenPorchSF
## Length:1460  Length:1460  Min.      : 0.00  Min.      : 0.00
## Class :character  Class :character  1st Qu.: 0.00  1st Qu.: 0.00
## Mode  :character  Mode  :character  Median : 0.00  Median : 25.00
##
##      Mean   : 94.24  Mean   : 46.66
##      3rd Qu.:168.00  3rd Qu.: 68.00
##      Max.   :857.00  Max.   :547.00
##
##      EnclosedPorch      3SsnPorch      ScreenPorch      PoolArea
## Min.      : 0.00  Min.      : 0.00  Min.      : 0.00  Min.      : 0.000
## 1st Qu.: 0.00  1st Qu.: 0.00  1st Qu.: 0.00  1st Qu.: 0.000
## Median : 0.00  Median : 0.00  Median : 0.00  Median : 0.000
## Mean   : 21.95  Mean   : 3.41  Mean   : 15.06  Mean   : 2.759
## 3rd Qu.: 0.00  3rd Qu.: 0.00  3rd Qu.: 0.00  3rd Qu.: 0.000
## Max.   :552.00  Max.   :508.00  Max.   :480.00  Max.   :738.000
##
##      PoolQC      Fence      MiscFeature      MiscVal
## Length:1460  Length:1460  Length:1460  Min.      : 0.00
## Class :character  Class :character  Class :character  1st Qu.: 0.00
## Mode  :character  Mode  :character  Mode  :character  Median : 0.00
##
##      Mean   : 43.49
##      3rd Qu.: 0.00

```



```
##                                     Max.      :15500.00
##
##      MoSold      YrSold      SaleType      SaleCondition
##  Min.   : 1.000   Min.    :2006   Length:1460   Length:1460
##  1st Qu.: 5.000   1st Qu.:2007   Class :character   Class :character
##  Median : 6.000   Median :2008   Mode  :character   Mode  :character
##  Mean   : 6.322   Mean    :2008
##  3rd Qu.: 8.000   3rd Qu.:2009
##  Max.   :12.000   Max.    :2010
##
##      SalePrice
##  Min.   : 34900
##  1st Qu.:129975
##  Median :163000
##  Mean   :180921
##  3rd Qu.:214000
##  Max.   :755000
##
```

```
# Checking To See Missing Data Values, and Splice the Code For Missing Values
sum(is.na(test))
```

```
## [1] 7000
```

```
sum(is.na(train))
```

```
## [1] 6965
```

```
NA_values_test=data.frame(NA_value=colSums(is.na(test)))
NA_values_train=data.frame(NA_value=colSums(is.na(train)))
head(NA_values_test,21)
```

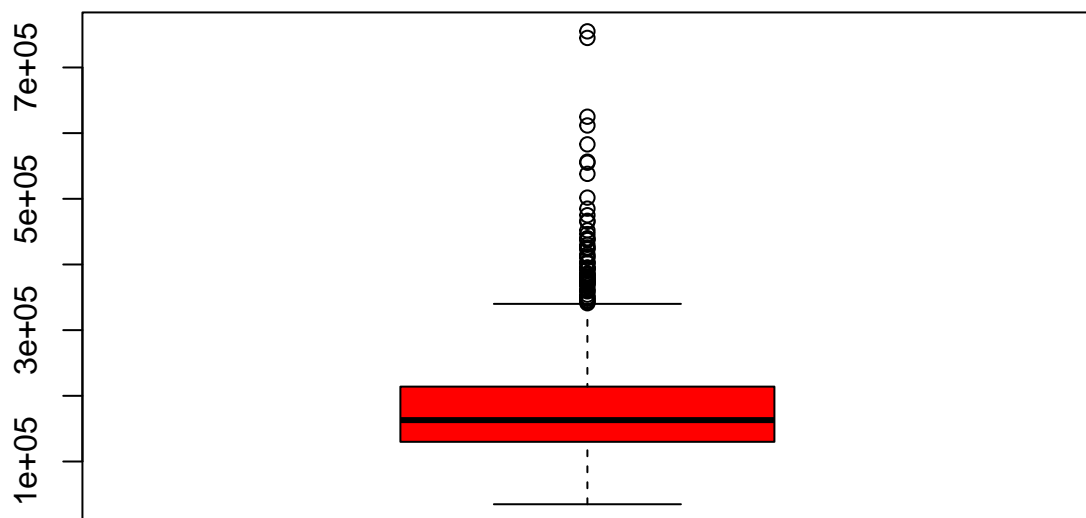
```
##      NA_value
## Id          0
## MSSubClass   0
## MSZoning     4
## LotFrontage 227
## LotArea      0
## Street       0
## Alley        1352
## LotShape     0
## LandContour  0
## Utilities    2
## LotConfig    0
## LandSlope    0
## Neighborhood 0
## Condition1   0
## Condition2   0
## BldgType     0
## HouseStyle   0
## OverallQual  0
## OverallCond  0
## YearBuilt    0
## YearRemodAdd 0
```

```
head(NA_values_train,21)
```

```
##           NA_value
## Id           0
## MSSubClass    0
## MSZoning      0
## LotFrontage  259
## LotArea       0
## Street        0
## Alley        1369
## LotShape      0
## LandContour   0
## Utilities     0
## LotConfig     0
## LandSlope     0
## Neighborhood  0
## Condition1    0
## Condition2    0
## BldgType      0
## HouseStyle    0
## OverallQual   0
## OverallCond   0
## YearBuilt     0
## YearRemodAdd  0
```

Boxplot and Histogram of sale price

```
# Box
boxplot(train$SalePrice, col = "red")
```



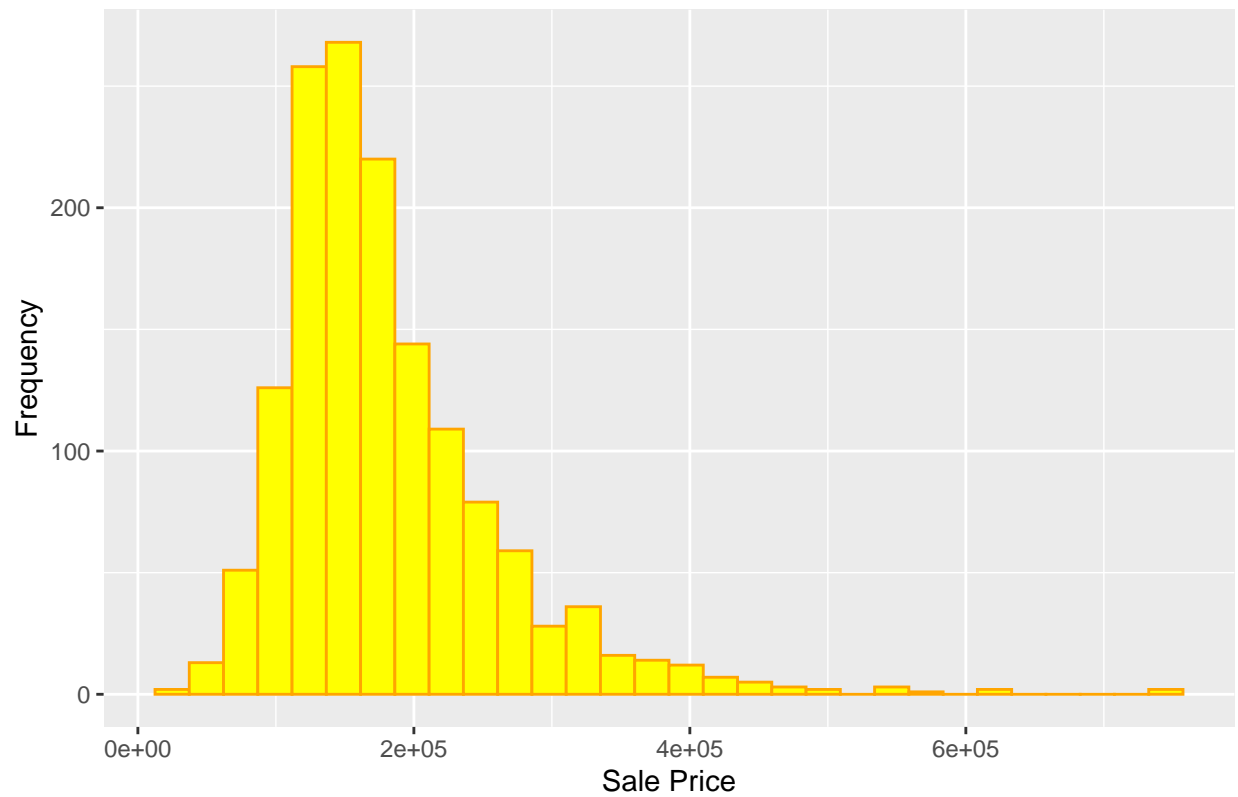
```
hist(train$SalePrice, col = "green")
```



```
# Frequency Sorted Histogram of Sale Prices
ggplot(train, aes(x = SalePrice)) +
  geom_histogram(fill = "yellow", color = "orange") +
  xlab("Sale Price") +
  ylab("Frequency") +
  ggtitle("Distribution of Sale Price")
```

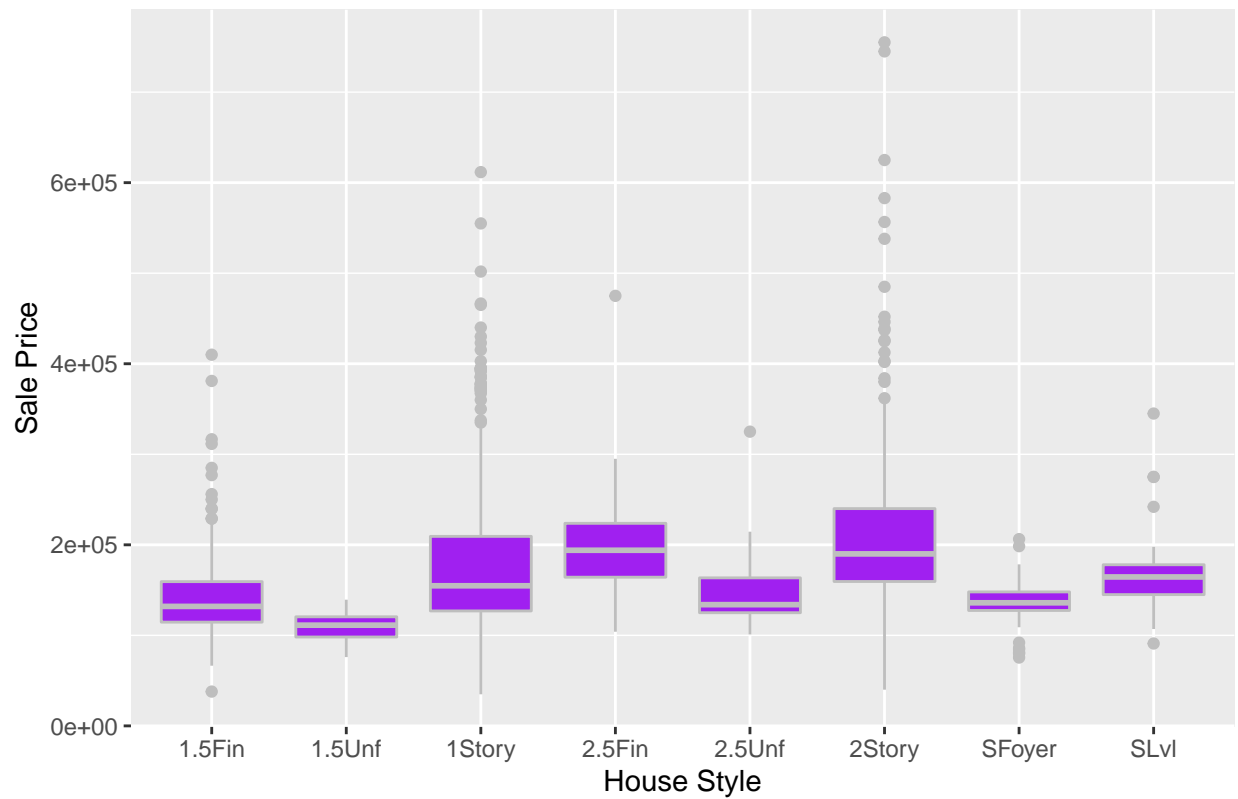
```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```

Distribution of Sale Price



```
# Style of Dwelling Relative to Sale Price  
ggplot(train, aes(x = HouseStyle, y = SalePrice)) +  
  geom_boxplot(fill = "purple", color = "grey") +  
  xlab("House Style") +  
  ylab("Sale Price") +  
  ggtitle("Sale Price by House Style")
```

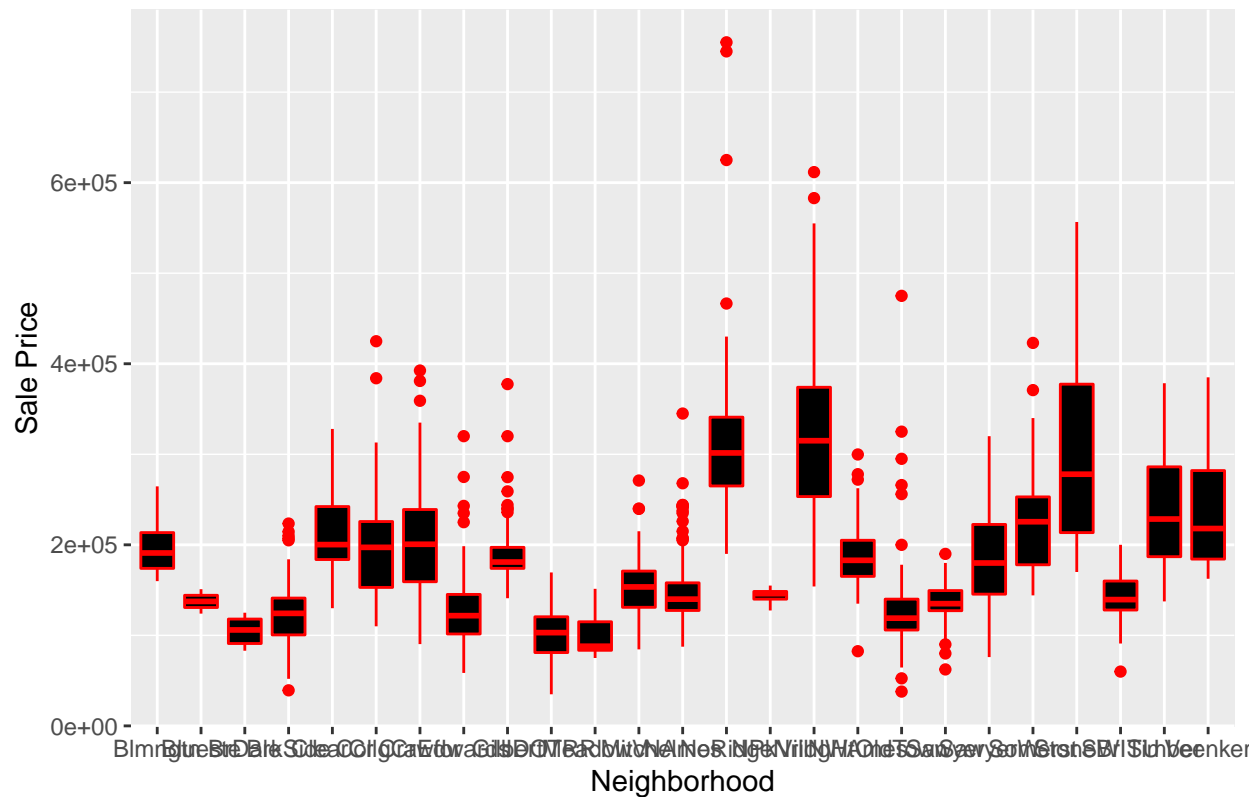
Sale Price by House Style



Neighborhood Box Plot

```
ggplot(train, aes(x = Neighborhood, y = SalePrice)) +
  geom_boxplot(fill = "black", color = "red") +
  xlab("Neighborhood") +
  ylab("Sale Price") +
  ggtitle("Sale Price by Neighborhood")
```

Sale Price by Neighborhood



Correlation Matrix

```
p_node <- train$SalePrice
```

```
# Percentage of Missing Values in Each Column
train <- train[,-81]
ncol(train)
```

```
## [1] 80
```

```
missing_percentages <- colMeans(is.na(train))
```

```
# Select Columns With Missing Values Less Than or Equal to 70%
retain_columns <- missing_percentages <= 0.30
```

```
# Keep The Selected Columns
train <- train[, retain_columns]
test <- test[, retain_columns]
train = cbind(train, SalePrice = p_node)
```

```
# Cleaning Up the ID Column
tidy_house_data <- train %>%
select(-Id)
```

```
test <- test %>%
select(-Id)
```

Assigning the Missing Values

```
tidy_house_data$LotFrontage[is.na(tidy_house_data$LotFrontage)] <- mean(tidy_house_data$LotFrontage, na.rm = TRUE)
tidy_house_data$MasVnrArea[is.na(tidy_house_data$MasVnrArea)] <- mean(tidy_house_data$MasVnrArea, na.rm = TRUE)
tidy_house_data$GarageYrBlt[is.na(tidy_house_data$GarageYrBlt)] <- mean(tidy_house_data$GarageYrBlt, na.rm = TRUE)
tidy_house_data$BsmtFinSF1[is.na(tidy_house_data$BsmtFinSF1)] <- mean(tidy_house_data$BsmtFinSF1, na.rm = TRUE)
tidy_house_data$BsmtFinSF2[is.na(tidy_house_data$BsmtFinSF2)] <- mean(tidy_house_data$BsmtFinSF2, na.rm = TRUE)
tidy_house_data$BsmtUnfSF[is.na(tidy_house_data$BsmtUnfSF)] <- mean(tidy_house_data$BsmtUnfSF, na.rm = TRUE)
tidy_house_data$TotalBsmtSF[is.na(tidy_house_data$TotalBsmtSF)] <- mean(tidy_house_data$TotalBsmtSF, na.rm = TRUE)
tidy_house_data$BsmtFullBath[is.na(tidy_house_data$BsmtFullBath)] <- mean(tidy_house_data$BsmtFullBath, na.rm = TRUE)
tidy_house_data$BsmtHalfBath[is.na(tidy_house_data$BsmtHalfBath)] <- mean(tidy_house_data$BsmtHalfBath, na.rm = TRUE)
tidy_house_data$GarageCars[is.na(tidy_house_data$GarageCars)] <- mean(tidy_house_data$GarageCars, na.rm = TRUE)
tidy_house_data$GarageArea[is.na(tidy_house_data$GarageArea)] <- mean(tidy_house_data$GarageArea, na.rm = TRUE)
```

```
Mode <- function(x) {
  ux <- unique(x)
  ux[which.max(tabulate(match(x, ux)))]
}
```

Replacing the Non-Existent Values With the Mode

```
for (i in 1:ncol(tidy_house_data)) {
  if (class(tidy_house_data[,i]) != "numeric" & sum(is.na(tidy_house_data[,i])) > 0) {
    tidy_house_data[,i] <- ifelse(is.na(tidy_house_data[,i]), Mode(tidy_house_data[,i]), tidy_house_data[,i])
  }
}
class(tidy_house_data$BsmtCond)
```

```
## [1] "character"
```

Factorizing Non-Numeric Variables

```
non_numeric_vars <- sapply(tidy_house_data, function(x) !is.numeric(x))
tidy_house_data[non_numeric_vars] <- lapply(tidy_house_data[non_numeric_vars], factor)
```

Selecting Numeric Variables

```
numeric_vars <- sapply(tidy_house_data, is.numeric)
data_numeric <- tidy_house_data[, numeric_vars]
tidy_house_data$GrLivArea
```

```
## [1] 1710 1262 1786 1717 2198 1362 1694 2090 1774 1077 1040 2324 912 1494
## [15] 1253 854 1004 1296 1114 1339 2376 1108 1795 1060 1060 1600 900 1704
## [29] 1600 520 1317 1228 1234 1700 1561 2452 1097 1297 1057 1152 1324 1328
## [43] 884 938 1150 1752 2149 1656 1452 955 1470 1176 816 1842 1360 1425
## [57] 1739 1720 2945 780 1158 1111 1370 1710 2034 2473 2207 1479 747 2287
## [71] 2223 845 1718 1086 1605 988 952 1285 1768 1230 2142 1337 1563 1065
## [85] 1474 2417 1560 1224 1526 990 1040 1235 964 2291 1786 1470 1588 960
## [99] 835 1225 1610 1732 1535 1226 1818 1992 1047 789 1517 1844 1855 1430
## [113] 2696 2259 2320 1458 1092 1125 3222 1456 988 1123 1080 1199 1586 754
## [127] 958 840 1348 1053 2157 2054 1327 1296 1721 1682 1214 1959 1852 1764
## [141] 864 1734 1385 1501 1728 1709 875 2035 1080 1344 969 1710 1993 1252
## [155] 1200 1096 1040 1968 1947 2462 1232 2668 1541 882 1616 1355 1867 2161
```



```

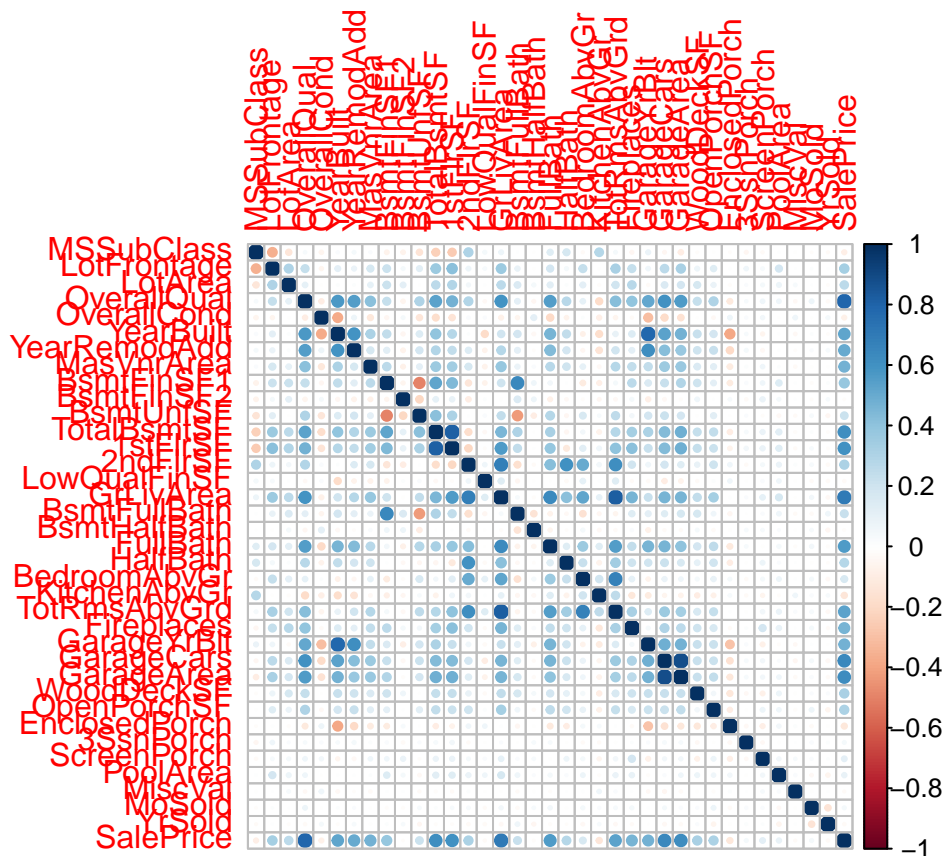
## [169] 1720 1707 1382 1656 1767 1362 1651 2158 2060 1920 2234 968 1525 1802
## [183] 1340 2082 1252 3608 1217 1656 1224 1593 2727 1479 1431 1709 864 1456
## [197] 1726 3112 2229 1713 1121 1279 1310 848 1284 1442 1696 1100 2062 1092
## [211] 864 1212 1852 990 1392 1236 1436 1328 1954 1248 1498 2267 1552 864
## [225] 2392 1302 2520 987 912 1555 1194 2794 987 894 1960 987 1414 1744
## [239] 1694 1487 1566 866 1440 1217 2110 1872 1928 1375 1668 2144 1306 1625
## [253] 1640 1302 1314 2291 1728 1604 1792 882 1382 2574 1212 1316 764 1422
## [267] 1511 2192 778 1113 1939 1363 2270 1632 816 1548 1560 864 2121 2022
## [281] 1982 1262 1314 1468 1575 1250 1734 858 900 1396 1919 1716 1716 2263
## [295] 1644 1003 1558 1950 1743 1152 1336 2452 1541 894 3493 2000 2243 1406
## [309] 861 1944 1501 972 1118 2036 1641 1432 2353 1959 2646 1472 2596 2468
## [323] 2730 1163 2978 803 1719 1383 2134 1192 1728 1056 1629 1358 1638 1786
## [337] 1922 1536 1621 1215 1908 841 1040 1684 1112 1577 958 1478 1626 2728
## [351] 1869 1453 1111 720 1595 1200 1167 1142 1352 1924 912 1505 1922 987
## [365] 1574 1344 1394 1431 1268 1287 1664 1588 752 1319 1928 904 914 2466
## [379] 1856 1800 1691 1301 1797 784 1953 1269 1184 1125 1479 2332 1367 1961
## [393] 882 788 1034 1144 894 1812 1077 1550 1288 1310 672 2263 1572 1620
## [407] 1639 1680 2172 2078 1276 1056 1478 1028 2097 1340 1400 2624 1134 1056
## [421] 1344 1602 988 2630 1196 1389 1644 907 1208 1412 987 1198 1365 1604
## [435] 630 1661 1118 904 694 1196 2402 1440 1573 1258 1908 1689 1888 1886
## [449] 1376 1183 813 1533 1756 1590 1728 1242 1344 1663 1666 1203 1935 1135
## [463] 864 1660 1040 1414 1277 1644 1634 1710 1502 1969 1072 1976 1652 970
## [477] 1493 2643 1718 1131 1850 1792 1826 1216 999 1113 1073 1484 2414 630
## [491] 1304 1578 1456 1269 886 720 3228 1820 899 912 1218 1768 1214 1801
## [505] 1322 1960 1911 1218 1378 1041 1363 1368 864 1080 789 2020 2119 2344
## [519] 1796 2080 1294 1244 1664 4676 2398 1266 928 2713 605 2515 1509 1362
## [533] 827 334 1414 1347 1724 864 1159 1601 1838 2285 1680 767 1496 2183
## [547] 1635 768 825 2094 1069 928 1717 1126 2046 1048 1092 1336 1446 1557
## [561] 1392 1389 996 1674 2295 1647 2504 1535 2132 943 1728 864 1692 1430
## [575] 1109 1216 1477 1320 1392 1795 1429 2042 816 2775 1573 2028 838 860
## [589] 1473 935 1582 2296 816 848 924 1826 1368 1402 1647 1556 1904 1375
## [603] 1915 1200 1494 1986 1040 2008 3194 1029 2153 1032 1872 1120 630 1054
## [617] 1509 832 1828 2262 864 2614 980 1512 1790 1116 1422 1520 2080 1350
## [631] 1750 1554 1411 1056 1056 3395 800 1387 796 1567 1518 1929 2704 1620
## [645] 1766 981 1048 1094 1839 630 1665 1510 1716 1469 2113 1092 1053 1502
## [659] 1458 1486 1935 2448 1392 1181 2097 1936 2380 1679 1437 1180 1476 1369
## [673] 1208 1839 1136 1441 1774 792 2046 988 923 1520 1291 1668 1839 2090
## [687] 1761 1102 1419 1362 848 4316 2519 1073 1539 1137 616 1148 894 1391
## [701] 1800 1164 2576 1812 1484 1092 1824 1324 1456 904 729 1178 1228 960
## [715] 1479 1350 2554 1178 2418 971 1742 848 864 1470 1698 864 1680 1232
## [729] 1776 1208 1616 1146 2031 1144 948 1768 1040 1801 1200 1728 1432 912
## [743] 1349 1464 1337 2715 2256 2640 1720 1529 1140 1320 1494 2098 1026 1471
## [757] 1768 1386 1501 2531 864 1301 1547 2365 1494 1506 1714 1750 1836 3279
## [771] 858 1220 1117 912 1973 1204 1614 894 2020 1004 1253 1603 1430 1110
## [785] 1484 1342 1652 2084 901 2087 1145 1062 2013 1496 1895 1564 1285 773
## [799] 3140 1768 1688 1196 1456 2822 1128 1428 980 1576 1086 2138 1309 848
## [813] 1044 1442 1250 1661 1008 1689 1052 1358 1640 936 1733 1489 1489 2084
## [827] 784 1434 2126 1223 1392 1200 1829 1516 1144 1067 1559 987 1099 1200
## [841] 1482 1539 1165 1800 1416 1701 1775 864 2358 1855 848 1456 1646 1445
## [855] 1779 1040 1026 1481 1370 2654 1426 1039 1097 1148 1372 1002 1646 1120
## [869] 2320 1949 894 1682 910 1268 1131 2610 1040 2224 1155 864 1090 1717
## [883] 1593 2230 892 1709 1712 1393 2217 1505 924 1683 1068 1383 1535 1796
## [897] 951 2240 2364 1236 858 1306 1509 1670 902 1063 1636 2057 902 1484
## [911] 2274 1268 1015 2002 1224 1092 480 1229 2127 1414 1721 2200 1316 1617

```

```
## [925] 1686 1126 2374 1978 1788 2236 1466 925 1905 1500 2069 747 1200 1971
## [939] 1962 2403 1728 2060 1440 1632 1344 1869 1144 1629 1776 1381 864 965
## [953] 768 1968 980 1958 1229 1057 1337 1416 858 2872 1548 1800 1894 1484
## [967] 1308 1098 968 1095 1192 1626 918 1428 2019 1382 869 1241 894 1121
## [981] 999 2612 1266 2290 1734 1164 1635 1940 2030 1576 2392 1742 1851 1500
## [995] 1718 1230 1050 1442 1077 1208 944 691 1574 1680 1504 985 1657 1092
## [1009] 1710 1522 1271 1664 1502 1022 1082 1665 1504 1360 1472 1506 1132 1220
## [1023] 1248 1504 2898 882 1264 1646 1376 1218 1928 3082 2520 1654 954 845
## [1037] 1620 2263 1344 630 1803 1632 1306 2329 2524 1733 2868 990 1771 930
## [1051] 1302 1316 1977 1526 1989 1523 1364 1850 2184 1991 1338 894 2337 1103
## [1065] 1154 2260 1571 1611 2521 893 1048 1556 1456 1426 1240 1740 1466 1096
## [1079] 848 990 1258 1040 1459 1251 1498 996 1092 1953 1709 1247 1040 1252
## [1093] 1694 1200 936 1314 1355 1088 1324 1601 438 950 1134 1194 1302 2622
## [1107] 1442 2021 1690 1836 1658 1964 816 1008 833 1734 1419 894 1601 1040
## [1121] 1012 1552 960 698 1482 1005 1555 1530 1959 936 1981 974 2210 2020
## [1135] 1600 986 1252 1020 1567 1167 952 1868 2828 1006 924 1576 1298 1564
## [1149] 1111 1482 932 1466 1811 816 1820 1437 1265 1314 1580 1876 1456 1640
## [1163] 894 1258 1432 1502 1694 1671 2108 3627 1118 1261 1250 3086 2345 2872
## [1177] 923 1224 1343 1124 2514 1652 4476 1130 1572 1221 1699 1624 1660 1804
## [1191] 1622 1441 1472 1224 1352 1456 1863 1690 1212 1382 864 1779 1348 1630
## [1205] 1074 2196 1056 1700 1283 1660 1845 1752 672 960 999 894 1902 1314
## [1219] 912 1218 912 1211 1846 2136 1490 1138 1933 912 1702 1507 2620 1190
## [1233] 1224 1188 1964 1784 1626 1948 1141 1484 1768 1689 1173 2076 1517 1868
## [1247] 1553 1034 2058 988 2110 1405 874 2167 1656 1367 1987 864 1166 1054
## [1261] 1675 1050 1788 1824 1337 1452 1889 2018 3447 1524 1524 1489 935 1357
## [1275] 1250 1920 1395 1724 2031 1128 1573 1339 1040 1824 2447 1412 1328 1582
## [1289] 1659 1970 1152 1302 2372 1664 864 1052 1128 1072 5642 1246 1983 1494
## [1303] 2526 1616 1708 1652 1368 990 1122 1294 1902 1274 2810 2599 948 2112
## [1317] 1630 1352 1787 948 1478 720 1923 708 1795 796 774 816 2792 1632
## [1331] 1588 954 816 1360 1365 1334 1656 693 1861 864 872 1114 2169 1913
## [1345] 1456 960 2156 1776 1494 2358 2634 1716 1176 3238 1865 1920 892 1078
## [1359] 1573 1980 2601 1530 1738 1412 1200 1674 1790 1475 848 1668 1374 1661
## [1373] 2097 2633 1958 1571 790 1604 987 1394 864 2117 1762 1416 1258 1154
## [1387] 2784 2526 1746 1218 1525 1584 900 1912 1500 2482 1687 1513 1904 1608
## [1401] 1158 1593 1294 1464 1214 1646 768 833 1363 2093 1840 1668 1040 1844
## [1415] 1848 1569 2290 2450 1144 1844 1416 1069 848 2201 1344 1252 2127 1558
## [1429] 804 1440 1838 958 968 1792 1126 1537 864 1932 1236 1725 2555 848
## [1443] 2007 952 1422 913 1188 2090 1346 630 1792 1578 1072 1140 1221 1647
## [1457] 2073 2340 1078 1256
```

```
# Identifying Categorical Variables
categorical_vars <- sapply(tidy_house_data, is.factor)

# Display the Correlation Matrix
cor_matrix <- cor(data_numeric)
corrplot(cor_matrix, method = "circle")
```



Scatterplots and Various Box Plots

It's interesting to note that in the cleaned house data, the `SalePrice` variable is strongly positively correlated with several other variables, including `OverallQual`, `GrLivArea`, `TotalBsmtSF`, `1stFlrSF`, `FullBath`, `GarageArea`, and `GarageCars`. This suggests that these variables may have a significant impact on the price of a house. For example, `OverallQual` represents the overall quality of the house, and it makes sense that a higher quality house would command a higher price. Similarly, `GrLivArea`, `TotalBsmtSF`, and `1stFlrSF` all measure the square footage of the house, which is likely to be an important factor in determining its value. Whereas the `FullBath` variable represents the number of full bathrooms in the house, which could also have a significant impact on its value. And finally, `GarageArea` and `GarageCars` both measure the size and capacity of the garage, which may be important to buyers who own vehicles or have other equipment that they need to store. These strong positive correlations suggest that these variables are important predictors of the `SalePrice` variable, and may be useful in developing a predictive model for house prices.

Herein, to demonstrate, scatterplots and boxplots will help visualize the relationship between the price and common predictors.

```
# Plotting Scatter Plots into a Grid
p1 = ggplot(data = tidy_house_data, aes(x = OverallQual, y = SalePrice)) + geom_jitter() + geom_smooth(linetype = 'solid', se = FALSE)

p2 = ggplot(data = tidy_house_data, aes(x = TotalBsmtSF, y = SalePrice)) + geom_jitter() + geom_smooth(linetype = 'solid', se = FALSE)

p3 = ggplot(data = tidy_house_data, aes(x = GrLivArea, y = SalePrice)) + geom_jitter() + geom_smooth(linetype = 'solid', se = FALSE)

p4 = ggplot(data = tidy_house_data, aes(x = GarageArea, y = SalePrice)) + geom_jitter() + geom_smooth(linetype = 'solid', se = FALSE)
```

```

p5 = ggplot(data = tidy_house_data, aes(x = GarageArea, y = TotalBsmtSF)) + geom_jitter() + geom_smooth()
p6 = ggplot(data = tidy_house_data, aes(x = GarageArea, y = `1stFlrSF`)) + geom_jitter() + geom_smooth()
p7 = ggplot(data = tidy_house_data, aes(x = OverallQual, y = TotalBsmtSF)) + geom_jitter() + geom_smooth()

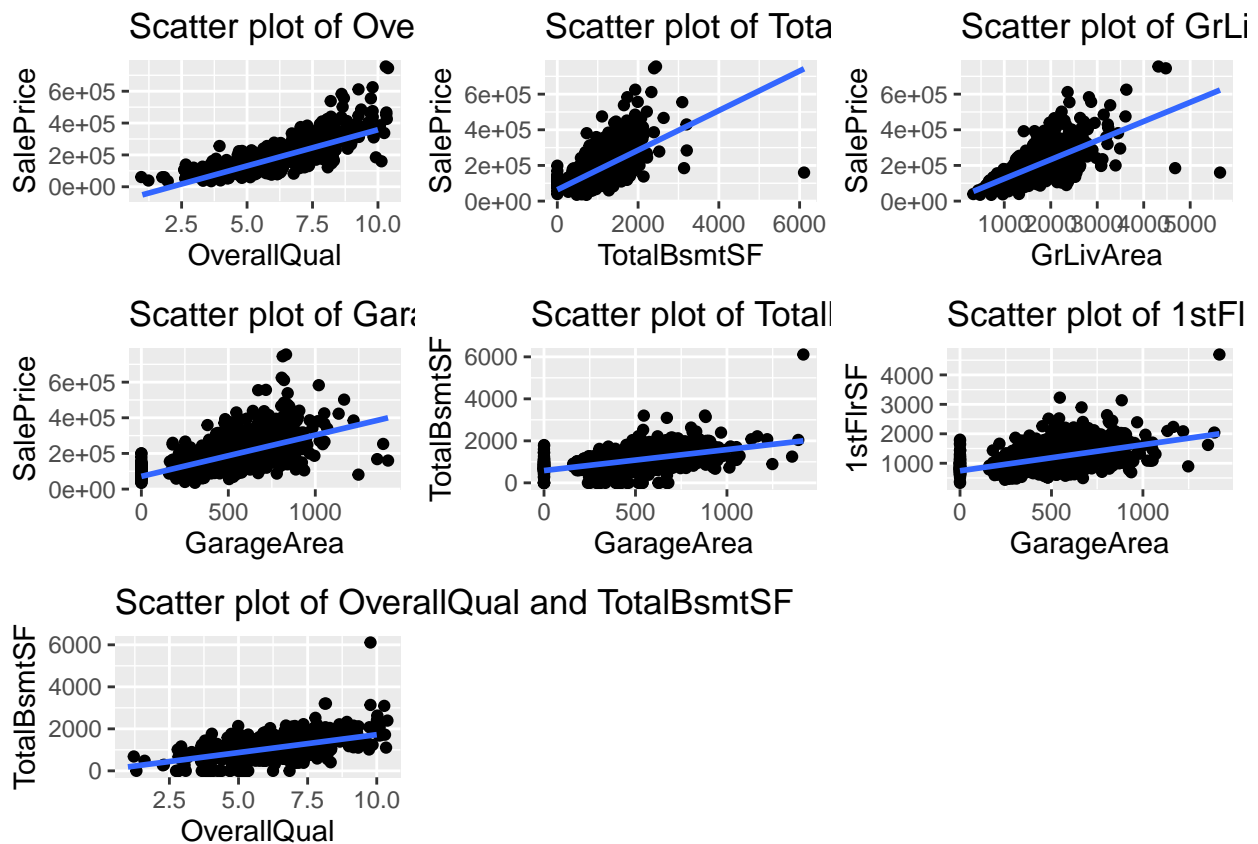
grid.arrange(p1, p2, p3, p4, p5, p6, p7, nrow=3)

```

```

## 'geom_smooth()' using formula 'y ~ x'
## 'geom_smooth()' using formula 'y ~ x'
## 'geom_smooth()' using formula 'y ~ x'
## 'geom_smooth()' using formula 'y ~ x'
## 'geom_smooth()' using formula 'y ~ x'
## 'geom_smooth()' using formula 'y ~ x'
## 'geom_smooth()' using formula 'y ~ x'

```



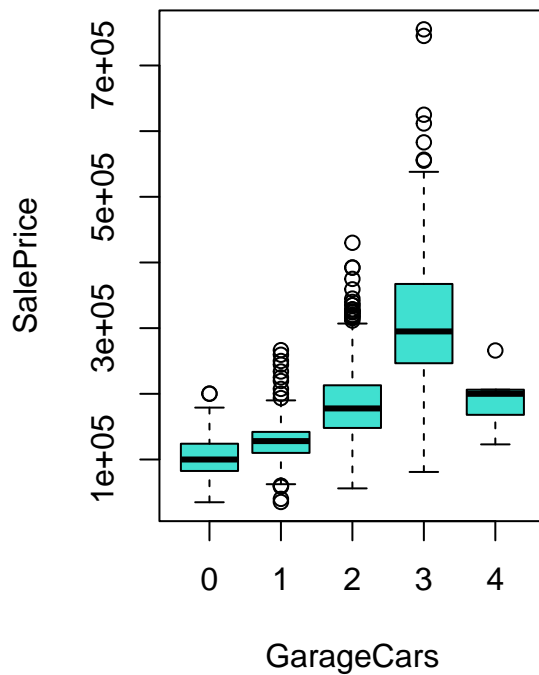
```

# Plotting Boxplots
par(mfrow=c(1, 2))

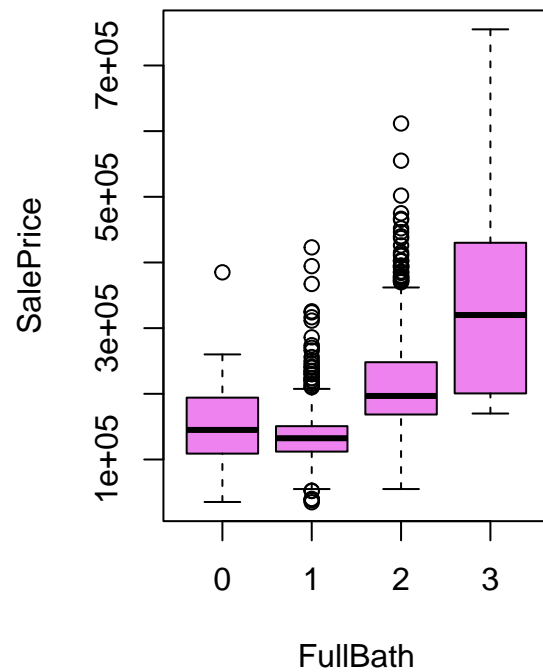
boxplot(SalePrice~GarageCars,data=tidy_house_data,main="Boxplot of GarageCars and SalePrice",col = "turquoise")
boxplot(SalePrice~FullBath,data=tidy_house_data,main="Boxplot of FullBath and SalePrice",col = "violet")

```

Boxplot of GarageCars and SalePr



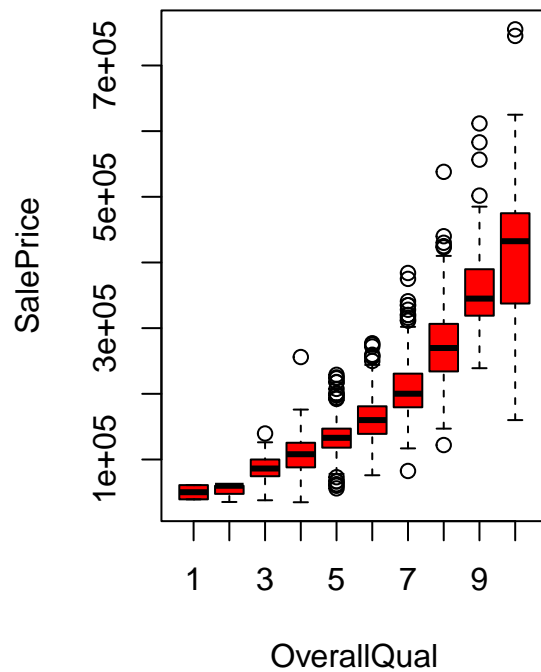
Boxplot of FullBath and SalePric



```
boxplot(SalePrice~OverallQual, data=tidy_house_data,main = "Boxplot of OverallQual and SalePrice", col = "cyan")

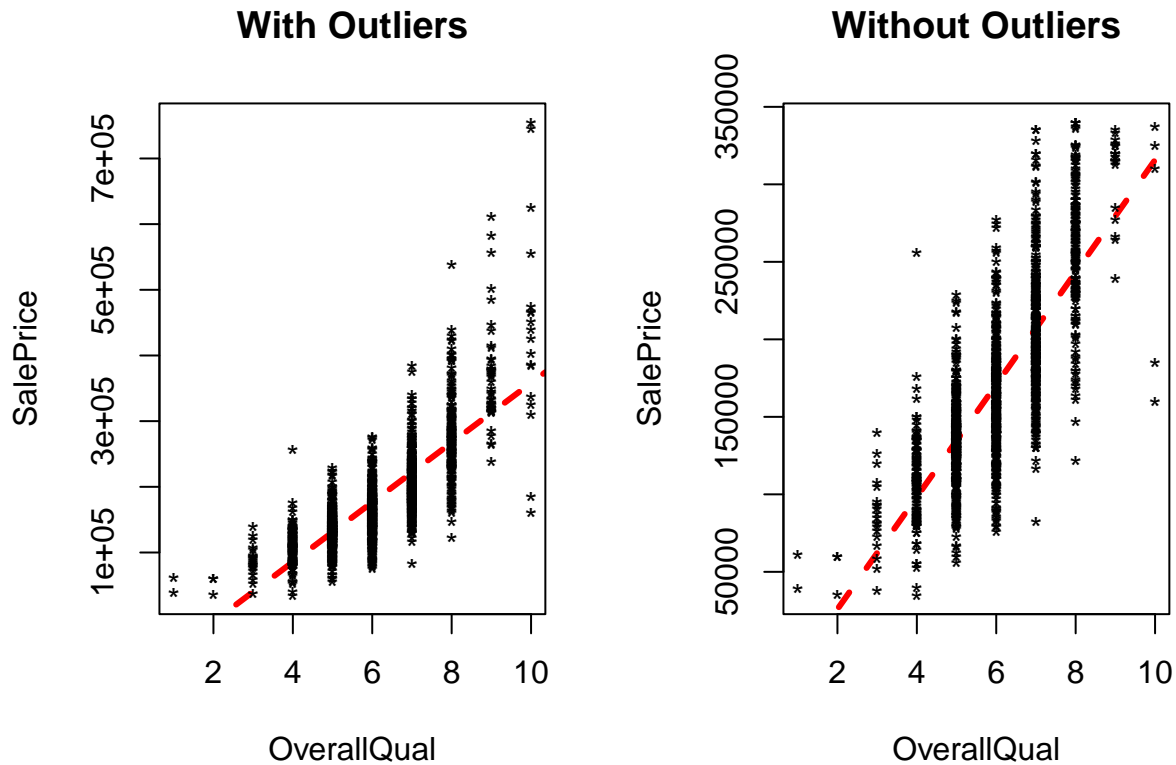
# Plotting Two Scatterplots, identifying SalePrice and OverallQual,
# but to compare two graphs with and without outliers
outliers=boxplot(tidy_house_data$SalePrice,plot=FALSE)$out
outliers_data=tidy_house_data[which(tidy_house_data$SalePrice %in% outliers),]
tidy_house_data1= tidy_house_data[-which(tidy_house_data$SalePrice %in% outliers),]
par(mfrow=c(1, 2))
```

Boxplot of OverallQual and SalePr



```
plot(tidy_house_data$OverallQual, tidy_house_data$SalePrice, main="With Outliers", xlab="OverallQual", ylab="SalePrice",
      abline(lm(SalePrice ~ OverallQual, data=tidy_house_data), col = "red", lwd=3, lty=2))

# Original Data Plot W/O Outliers.
plot(tidy_house_data1$OverallQual, tidy_house_data1$SalePrice, main="Without Outliers", xlab="OverallQual", ylab="SalePrice",
      abline(lm(SalePrice ~ OverallQual, data=tidy_house_data1), col = "red", lwd=3, lty=2))
```



Modeling Training Data

```
model.full <- lm (formula = SalePrice ~ ., data = tidy_house_data1)
summary(model.full)
```

```
##
## Call:
## lm(formula = SalePrice ~ ., data = tidy_house_data1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -115885   -7525        0    7708   69759
##
## Coefficients: (4 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -4.667e+05  7.580e+05  -0.616  0.538179
## MSSubClass     -9.959e+00  5.844e+01  -0.170  0.864725
## MSZoningFV      3.426e+04  8.507e+03   4.027  6.01e-05 ***
## MSZoningRH      3.138e+04  8.393e+03   3.739  0.000194 ***
## MSZoningRL      3.279e+04  7.180e+03   4.567  5.47e-06 ***
## MSZoningRM      2.252e+04  6.712e+03   3.355  0.000820 ***
## LotFrontage     3.572e+01  3.185e+01   1.122  0.262281
## LotArea         7.959e-01  9.293e-02   8.564  < 2e-16 ***
```

## StreetPave	3.312e+04	8.690e+03	3.811	0.000146	***
## LotShapeIR2	4.129e+00	3.193e+03	0.001	0.998968	
## LotShapeIR3	4.148e+03	6.425e+03	0.646	0.518692	
## LotShapeReg	5.017e+01	1.165e+03	0.043	0.965663	
## LandContourHLS	-5.443e+02	3.825e+03	-0.142	0.886858	
## LandContourLow	-6.863e+03	4.558e+03	-1.506	0.132385	
## LandContourLvl	1.432e+03	2.667e+03	0.537	0.591416	
## UtilitiesNoSeWa	-2.870e+04	1.864e+04	-1.540	0.123808	
## LotConfigCulDSac	5.144e+03	2.403e+03	2.141	0.032465	*
## LotConfigFR2	-5.865e+03	2.878e+03	-2.038	0.041806	*
## LotConfigFR3	-1.032e+04	8.921e+03	-1.156	0.247825	
## LotConfigInside	-1.025e+03	1.289e+03	-0.795	0.426762	
## LandSlopeMod	7.214e+03	2.903e+03	2.485	0.013095	*
## LandSlopeSev	-4.447e+04	8.158e+03	-5.452	6.08e-08	***
## NeighborhoodBlueste	2.251e+03	1.340e+04	0.168	0.866616	
## NeighborhoodBrDale	1.036e+04	7.806e+03	1.327	0.184931	
## NeighborhoodBrkSide	6.332e+03	6.699e+03	0.945	0.344721	
## NeighborhoodClearCr	-2.600e+03	6.601e+03	-0.394	0.693689	
## NeighborhoodCollgCr	-5.930e+03	5.160e+03	-1.149	0.250696	
## NeighborhoodCrawfor	1.716e+04	6.113e+03	2.807	0.005081	**
## NeighborhoodEdwards	-9.986e+03	5.692e+03	-1.754	0.079608	.
## NeighborhoodGilbert	-7.177e+03	5.500e+03	-1.305	0.192221	
## NeighborhoodIDOTRR	3.317e+03	7.617e+03	0.436	0.663239	
## NeighborhoodMeadowV	-9.165e+03	8.069e+03	-1.136	0.256255	
## NeighborhoodMitchel	-1.300e+04	5.796e+03	-2.243	0.025090	*
## NeighborhoodNames	-9.377e+03	5.559e+03	-1.687	0.091894	.
## NeighborhoodNoRidge	1.780e+04	6.159e+03	2.890	0.003930	**
## NeighborhoodNPkVill	2.944e+03	9.912e+03	0.297	0.766513	
## NeighborhoodNridgHt	1.428e+04	5.510e+03	2.592	0.009672	**
## NeighborhoodNWAmes	-1.027e+04	5.711e+03	-1.798	0.072380	.
## NeighborhoodOldTown	-1.781e+03	6.810e+03	-0.262	0.793706	
## NeighborhoodSawyer	-5.443e+03	5.779e+03	-0.942	0.346433	
## NeighborhoodSawyerW	-2.636e+03	5.515e+03	-0.478	0.632797	
## NeighborhoodSomerst	7.423e+03	6.491e+03	1.144	0.253035	
## NeighborhoodStoneBr	1.540e+04	6.537e+03	2.357	0.018604	*
## NeighborhoodSWISU	-2.583e+03	6.891e+03	-0.375	0.707863	
## NeighborhoodTimber	1.700e+03	5.845e+03	0.291	0.771170	
## NeighborhoodVeenker	8.162e+03	7.660e+03	1.065	0.286880	
## Condition1Feedr	5.386e+03	3.511e+03	1.534	0.125249	
## Condition1Norm	1.337e+04	2.922e+03	4.576	5.24e-06	***
## Condition1PosA	1.012e+04	7.054e+03	1.434	0.151840	
## Condition1PosN	1.412e+04	5.251e+03	2.688	0.007280	**
## Condition1RRAE	-8.776e+03	6.410e+03	-1.369	0.171218	
## Condition1RRAN	4.414e+03	4.870e+03	0.906	0.364932	
## Condition1RRNe	-7.791e+02	1.236e+04	-0.063	0.949737	
## Condition1RRNn	1.895e+04	9.101e+03	2.082	0.037551	*
## Condition2Feedr	4.993e+03	1.648e+04	0.303	0.762017	
## Condition2Norm	8.550e+03	1.427e+04	0.599	0.549334	
## Condition2PosA	6.909e+04	2.631e+04	2.625	0.008765	**
## Condition2PosN	-3.401e+05	2.378e+04	-14.301	< 2e-16	***
## Condition2RRAE	-1.072e+05	3.270e+04	-3.279	0.001072	**
## Condition2RRAN	-7.945e+03	2.223e+04	-0.357	0.720917	
## Condition2RRNn	1.862e+04	1.911e+04	0.974	0.330162	
## BldgType2fmCon	-7.489e+03	8.842e+03	-0.847	0.397170	

## BldgTypeDuplex	-8.551e+03	5.182e+03	-1.650	0.099188	.
## BldgTypeTwnhs	-1.892e+04	7.063e+03	-2.679	0.007487	**
## BldgTypeTwnhsE	-1.079e+04	6.363e+03	-1.695	0.090303	.
## HouseStyle1.5Unf	9.775e+03	5.589e+03	1.749	0.080551	.
## HouseStyle1Story	5.914e+03	3.151e+03	1.877	0.060825	.
## HouseStyle2.5Fin	-2.064e+04	8.728e+03	-2.365	0.018176	*
## HouseStyle2.5Unf	-6.327e+03	6.535e+03	-0.968	0.333176	
## HouseStyle2Story	-2.445e+03	2.539e+03	-0.963	0.335786	
## HouseStyleSFoyer	3.403e+03	4.428e+03	0.769	0.442330	
## HouseStyleSLvl	2.198e+03	3.950e+03	0.557	0.577935	
## OverallQual	6.391e+03	7.293e+02	8.763	< 2e-16	***
## OverallCond	5.083e+03	6.236e+02	8.151	9.21e-16	***
## YearBuilt	2.993e+02	5.405e+01	5.538	3.78e-08	***
## YearRemodAdd	1.041e+02	3.904e+01	2.666	0.007792	**
## RoofStyleGable	-4.570e+03	1.300e+04	-0.351	0.725324	
## RoofStyleGambrel	-1.851e+03	1.422e+04	-0.130	0.896443	
## RoofStyleHip	-4.575e+03	1.305e+04	-0.351	0.725928	
## RoofStyleMansard	8.283e+03	1.513e+04	0.547	0.584166	
## RoofStyleShed	8.223e+04	2.448e+04	3.360	0.000806	***
## RoofMatlCompShg	5.302e+05	2.500e+04	21.205	< 2e-16	***
## RoofMatlMembran	6.086e+05	3.495e+04	17.410	< 2e-16	***
## RoofMatlMetal	5.760e+05	3.433e+04	16.776	< 2e-16	***
## RoofMatlRoll	5.223e+05	3.055e+04	17.101	< 2e-16	***
## RoofMatlTar&Grv	5.298e+05	2.820e+04	18.787	< 2e-16	***
## RoofMatlWdShake	5.205e+05	2.736e+04	19.028	< 2e-16	***
## RoofMatlWdShngl	5.389e+05	2.670e+04	20.183	< 2e-16	***
## Exterior1stAsphShn	-8.887e+03	2.315e+04	-0.384	0.701065	
## Exterior1stBrkComm	-2.058e+04	1.956e+04	-1.052	0.293077	
## Exterior1stBrkFace	4.973e+03	8.973e+03	0.554	0.579497	
## Exterior1stCBlock	-2.399e+04	1.917e+04	-1.252	0.210914	
## Exterior1stCemntBd	-2.277e+04	1.933e+04	-1.178	0.239064	
## Exterior1stHdBoard	-1.533e+04	9.090e+03	-1.686	0.092080	.
## Exterior1stImStucc	-1.380e+04	1.975e+04	-0.699	0.484974	
## Exterior1stMetalSd	-7.877e+03	1.027e+04	-0.767	0.443052	
## Exterior1stPlywood	-1.090e+04	8.984e+03	-1.213	0.225435	
## Exterior1stStone	-3.868e+03	1.692e+04	-0.229	0.819254	
## Exterior1stStucco	-1.190e+04	1.006e+04	-1.182	0.237261	
## Exterior1stVinylSd	-1.454e+04	9.326e+03	-1.559	0.119287	
## Exterior1stWd Sdng	-1.940e+04	8.708e+03	-2.228	0.026082	*
## Exterior1stWdShing	-1.077e+04	9.422e+03	-1.143	0.253255	
## Exterior2ndAsphShn	3.505e+03	1.562e+04	0.224	0.822544	
## Exterior2ndBrk Cmn	1.025e+04	1.418e+04	0.723	0.469957	
## Exterior2ndBrkFace	1.943e+03	9.301e+03	0.209	0.834579	
## Exterior2ndCBlock	NA	NA	NA	NA	
## Exterior2ndCmentBd	2.600e+04	1.931e+04	1.347	0.178330	
## Exterior2ndHdBoard	6.954e+03	8.729e+03	0.797	0.425845	
## Exterior2ndImStucc	5.734e+03	1.051e+04	0.546	0.585446	
## Exterior2ndMetalSd	4.605e+03	9.990e+03	0.461	0.644904	
## Exterior2ndOther	3.280e+04	1.950e+04	1.682	0.092895	.
## Exterior2ndPlywood	3.238e+03	8.483e+03	0.382	0.702796	
## Exterior2ndStone	-5.418e+03	1.198e+04	-0.452	0.651188	
## Exterior2ndStucco	1.067e+04	9.641e+03	1.107	0.268658	
## Exterior2ndVinylSd	1.232e+04	9.010e+03	1.367	0.171779	
## Exterior2ndWd Sdng	1.270e+04	8.419e+03	1.508	0.131829	

## Exterior2ndWd Shng	5.281e+03	8.777e+03	0.602	0.547543	
## MasVnrTypeBrkFace	7.833e+03	4.798e+03	1.633	0.102831	
## MasVnrTypeNone	7.273e+03	4.845e+03	1.501	0.133584	
## MasVnrTypeStone	1.304e+04	5.129e+03	2.543	0.011125	*
## MasVnrArea	5.809e+00	4.473e+00	1.299	0.194357	
## ExterQualFa	5.637e+03	8.271e+03	0.682	0.495672	
## ExterQualGd	-3.730e+03	4.407e+03	-0.846	0.397500	
## ExterQualTA	-7.242e+03	4.653e+03	-1.556	0.119884	
## ExterCondFa	-1.769e+03	1.268e+04	-0.139	0.889099	
## ExterCondGd	-6.491e+03	1.211e+04	-0.536	0.592059	
## ExterCondPo	1.041e+04	2.213e+04	0.470	0.638205	
## ExterCondTA	-2.494e+03	1.208e+04	-0.206	0.836487	
## FoundationCBlock	2.145e+03	2.255e+03	0.951	0.341828	
## FoundationPConc	3.112e+03	2.424e+03	1.284	0.199482	
## FoundationSlab	2.557e+03	5.516e+03	0.464	0.643068	
## FoundationStone	6.259e+03	7.787e+03	0.804	0.421650	
## FoundationWood	-2.557e+04	1.044e+04	-2.450	0.014437	*
## BsmtQualFa	-9.589e+03	4.527e+03	-2.118	0.034361	*
## BsmtQualGd	-1.229e+04	2.527e+03	-4.862	1.32e-06	***
## BsmtQualTA	-1.190e+04	3.047e+03	-3.906	9.94e-05	***
## BsmtCondGd	4.078e+03	3.762e+03	1.084	0.278557	
## BsmtCondPo	7.426e+04	2.110e+04	3.520	0.000448	***
## BsmtCondTA	2.890e+03	2.993e+03	0.966	0.334418	
## BsmtExposureGd	7.903e+03	2.305e+03	3.429	0.000626	***
## BsmtExposureMn	-7.761e+02	2.214e+03	-0.351	0.725936	
## BsmtExposureNo	-2.446e+03	1.588e+03	-1.540	0.123840	
## BsmtFinType1BLQ	-2.005e+02	1.957e+03	-0.102	0.918399	
## BsmtFinType1GLQ	4.965e+03	1.786e+03	2.781	0.005510	**
## BsmtFinType1LwQ	-4.251e+03	2.649e+03	-1.605	0.108767	
## BsmtFinType1Rec	-1.495e+03	2.114e+03	-0.707	0.479443	
## BsmtFinType1Unf	6.848e+02	2.088e+03	0.328	0.742976	
## BsmtFinSF1	2.559e+01	3.442e+00	7.436	2.00e-13	***
## BsmtFinType2BLQ	-1.019e+04	5.426e+03	-1.878	0.060625	.
## BsmtFinType2GLQ	4.367e+02	6.644e+03	0.066	0.947601	
## BsmtFinType2LwQ	-1.041e+04	5.325e+03	-1.955	0.050779	.
## BsmtFinType2Rec	-7.278e+03	5.142e+03	-1.416	0.157173	
## BsmtFinType2Unf	-5.208e+03	5.449e+03	-0.956	0.339379	
## BsmtFinSF2	2.313e+01	6.089e+00	3.799	0.000153	***
## BsmtUnfSF	1.365e+01	2.871e+00	4.755	2.24e-06	***
## TotalBsmtSF	NA	NA	NA	NA	
## HeatingGasA	2.424e+03	1.737e+04	0.140	0.889051	
## HeatingGasW	2.427e+02	1.802e+04	0.013	0.989257	
## HeatingGrav	-6.776e+02	1.910e+04	-0.035	0.971699	
## HeatingOthW	-1.405e+04	2.167e+04	-0.648	0.516903	
## HeatingWall	1.667e+04	2.019e+04	0.825	0.409321	
## HeatingQCFA	4.508e+02	3.292e+03	0.137	0.891097	
## HeatingQCGd	-2.261e+03	1.469e+03	-1.539	0.124181	
## HeatingQCPo	-3.707e+03	1.880e+04	-0.197	0.843727	
## HeatingQCTA	-2.682e+03	1.461e+03	-1.836	0.066653	.
## CentralAirY	1.960e+03	2.734e+03	0.717	0.473528	
## ElectricalFuseF	1.246e+03	4.070e+03	0.306	0.759573	
## ElectricalFuseP	-2.909e+03	1.303e+04	-0.223	0.823409	
## ElectricalMix	-4.421e+04	3.147e+04	-1.405	0.160367	
## ElectricalSBrkr	-8.161e+02	2.081e+03	-0.392	0.695056	

## '1stFlrSF'	4.639e+01	3.839e+00	12.083	< 2e-16	***
## '2ndFlrSF'	5.738e+01	4.303e+00	13.335	< 2e-16	***
## LowQualFinSF	2.621e+01	1.294e+01	2.025	0.043075	*
## GrLivArea	NA	NA	NA	NA	
## BsmtFullBath	2.647e+03	1.409e+03	1.878	0.060617	.
## BsmtHalfBath	-1.501e+03	2.154e+03	-0.697	0.485897	
## FullBath	9.313e+02	1.599e+03	0.583	0.560330	
## HalfBath	-4.523e+02	1.525e+03	-0.297	0.766883	
## BedroomAbvGr	-7.970e+02	9.924e+02	-0.803	0.422101	
## KitchenAbvGr	-7.259e+03	4.009e+03	-1.810	0.070481	.
## KitchenQualFa	-1.750e+04	4.483e+03	-3.903	0.000100	***
## KitchenQualGd	-1.899e+04	2.670e+03	-7.111	2.00e-12	***
## KitchenQualTA	-1.888e+04	2.935e+03	-6.435	1.80e-10	***
## TotRmsAbvGrd	-2.165e+02	6.947e+02	-0.312	0.755330	
## FunctionalMaj2	-8.271e+03	1.019e+04	-0.811	0.417309	
## FunctionalMin1	1.039e+04	6.075e+03	1.710	0.087446	.
## FunctionalMin2	1.164e+04	6.108e+03	1.906	0.056907	.
## FunctionalMod	-2.385e+03	7.411e+03	-0.322	0.747624	
## FunctionalSev	-4.713e+04	2.049e+04	-2.300	0.021616	*
## FunctionalTyp	2.034e+04	5.286e+03	3.848	0.000126	***
## Fireplaces	2.853e+03	9.672e+02	2.949	0.003247	**
## GarageTypeAttchd	2.062e+04	7.728e+03	2.668	0.007727	**
## GarageTypeBasement	1.830e+04	8.960e+03	2.042	0.041349	*
## GarageTypeBuiltIn	2.016e+04	8.100e+03	2.489	0.012940	*
## GarageTypeCarPort	2.214e+04	1.025e+04	2.160	0.030964	*
## GarageTypeDetchd	2.036e+04	7.694e+03	2.646	0.008252	**
## GarageYrBlt	-9.256e+00	4.021e+01	-0.230	0.817988	
## GarageFinishRfn	-1.735e+03	1.435e+03	-1.210	0.226694	
## GarageFinishUnf	-3.530e+02	1.718e+03	-0.205	0.837221	
## GarageCars	4.174e+03	1.595e+03	2.617	0.008988	**
## GarageArea	1.328e+01	5.794e+00	2.293	0.022047	*
## GarageQualFa	-1.235e+04	1.256e+04	-0.983	0.325837	
## GarageQualGd	2.958e+03	1.303e+04	0.227	0.820454	
## GarageQualPo	-2.236e+04	2.069e+04	-1.081	0.279958	
## GarageQualTA	-6.743e+03	1.198e+04	-0.563	0.573570	
## GarageCondFa	-4.002e+03	3.822e+03	-1.047	0.295245	
## GarageCondGd	-2.195e+03	6.310e+03	-0.348	0.728013	
## GarageCondPo	-1.476e+03	9.758e+03	-0.151	0.879829	
## GarageCondTA	NA	NA	NA	NA	
## PavedDriveP	-5.502e+03	3.849e+03	-1.429	0.153171	
## PavedDriveY	-4.969e+02	2.413e+03	-0.206	0.836880	
## WoodDeckSF	1.575e+01	4.244e+00	3.710	0.000217	***
## OpenPorchSF	2.416e+01	8.239e+00	2.933	0.003428	**
## EnclosedPorch	8.607e+00	8.923e+00	0.965	0.334936	
## '3SsnPorch'	2.614e+01	1.662e+01	1.573	0.115956	
## ScreenPorch	3.127e+01	8.884e+00	3.520	0.000448	***
## PoolArea	2.614e+01	1.381e+01	1.893	0.058604	.
## MiscVal	6.843e-01	9.991e-01	0.685	0.493527	
## MoSold	5.585e+01	1.762e+02	0.317	0.751315	
## YrSold	-4.650e+02	3.705e+02	-1.255	0.209649	
## SaleTypeCon	3.136e+04	1.243e+04	2.522	0.011791	*
## SaleTypeConLD	1.316e+04	6.826e+03	1.929	0.054011	.
## SaleTypeConLI	-3.298e+03	8.878e+03	-0.371	0.710393	
## SaleTypeConLw	1.921e+03	8.565e+03	0.224	0.822613	

```
## SaleTypeCWD          2.004e+04  9.096e+03  2.203 0.027769 *
## SaleTypeNew          1.148e+04  1.095e+04  1.048 0.294864
## SaleTypeOth          8.097e+03  1.013e+04  0.799 0.424382
## SaleTypeWD          -6.233e+02  2.948e+03 -0.211 0.832613
## SaleConditionAdjLand  1.242e+04  1.024e+04  1.214 0.225163
## SaleConditionAlloca   6.861e+03  6.339e+03  1.082 0.279317
## SaleConditionFamily   2.518e+03  4.286e+03  0.587 0.557087
## SaleConditionNormal   7.771e+03  2.036e+03  3.817 0.000142 ***
## SaleConditionPartial -1.297e+03  1.051e+04 -0.123 0.901764
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16010 on 1170 degrees of freedom
## Multiple R-squared:  0.9389, Adjusted R-squared:  0.927
## F-statistic: 78.84 on 228 and 1170 DF,  p-value: < 2.2e-16
```

```
model0 <- lm(SalePrice ~ GarageArea + GarageCars + Functional + KitchenAbvGr + KitchenQual + `2ndFlrSF` +
summary(model0)
```

```
##
## Call:
## lm(formula = SalePrice ~ GarageArea + GarageCars + Functional +
##      KitchenAbvGr + KitchenQual + `2ndFlrSF` + `1stFlrSF` + BsmtUnfSF +
##      BsmtFinSF2 + BsmtFinSF1 + BsmtQual + RoofMatl + RoofStyle +
##      OverallQual + YearBuilt, data = tidy_house_data1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -322237  -12664       -1   12002  112283
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -8.997e+05  7.668e+04 -11.733  < 2e-16 ***
## GarageArea    2.427e+01  6.830e+00   3.554 0.000393 ***
## GarageCars    2.938e+03  1.994e+03   1.474 0.140753
## FunctionalMaj2 1.783e+03  1.237e+04   0.144 0.885366
## FunctionalMin1 1.467e+04  7.635e+03   1.921 0.054904 .
## FunctionalMin2 1.774e+04  7.530e+03   2.356 0.018628 *
## FunctionalMod   5.146e+03  9.072e+03   0.567 0.570613
## FunctionalSev  -7.649e+04  2.556e+04  -2.993 0.002811 **
## FunctionalTyp   2.716e+04  6.421e+03   4.230 2.50e-05 ***
## KitchenAbvGr  -2.292e+04  3.093e+03  -7.410 2.21e-13 ***
## KitchenQualFa  -2.622e+04  5.420e+03  -4.839 1.46e-06 ***
## KitchenQualGd  -1.636e+04  3.421e+03  -4.782 1.92e-06 ***
## KitchenQualTA  -2.978e+04  3.645e+03  -8.171 6.92e-16 ***
## `2ndFlrSF`     4.949e+01  1.939e+00  25.522  < 2e-16 ***
## `1stFlrSF`     5.739e+01  3.405e+00  16.856  < 2e-16 ***
## BsmtUnfSF       1.067e+01  2.962e+00   3.602 0.000328 ***
## BsmtFinSF2      1.940e+01  4.772e+00   4.065 5.08e-05 ***
## BsmtFinSF1      2.736e+01  3.078e+00   8.888  < 2e-16 ***
## BsmtQualFa     -2.077e+04  5.673e+03  -3.662 0.000260 ***
## BsmtQualGd     -1.607e+04  3.168e+03  -5.074 4.43e-07 ***
## BsmtQualTA     -2.131e+04  3.660e+03  -5.821 7.28e-09 ***
## RoofMatlCompShg  5.055e+05  2.657e+04  19.022  < 2e-16 ***
```

```
## RoofMatlMembran 5.542e+05 4.023e+04 13.777 < 2e-16 ***
## RoofMatlMetal 5.043e+05 3.942e+04 12.794 < 2e-16 ***
## RoofMatlRoll 5.035e+05 3.556e+04 14.158 < 2e-16 ***
## RoofMatlTar&Grv 4.839e+05 3.155e+04 15.336 < 2e-16 ***
## RoofMatlWdShake 4.996e+05 2.909e+04 17.174 < 2e-16 ***
## RoofMatlWdShngl 5.217e+05 2.950e+04 17.687 < 2e-16 ***
## RoofStyleGable -3.396e+04 1.719e+04 -1.976 0.048363 *
## RoofStyleGambrel -2.756e+04 1.865e+04 -1.478 0.139626
## RoofStyleHip -3.292e+04 1.726e+04 -1.907 0.056670 .
## RoofStyleMansard -2.758e+04 1.965e+04 -1.403 0.160797
## RoofStyleShed -1.792e+04 2.464e+04 -0.727 0.467293
## OverallQual 1.061e+04 8.476e+02 12.516 < 2e-16 ***
## YearBuilt 2.322e+02 3.484e+01 6.666 3.80e-11 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23390 on 1364 degrees of freedom
## Multiple R-squared: 0.8479, Adjusted R-squared: 0.8441
## F-statistic: 223.6 on 34 and 1364 DF, p-value: < 2.2e-16
```

Testing Data Predictions

```
# Dataframe Form of the Test Data
test <- as.data.frame(test)

test$LotFrontage[is.na(test$LotFrontage)] <- mean(test$LotFrontage, na.rm = TRUE)
test$MasVnrArea[is.na(test$MasVnrArea)] <- mean(test$MasVnrArea, na.rm = TRUE)
test$GarageYrBlt[is.na(test$GarageYrBlt)] <- mean(test$GarageYrBlt, na.rm = TRUE)
test$BsmtFinSF1[is.na(test$BsmtFinSF1)] <- mean(test$BsmtFinSF1, na.rm = TRUE)
test$BsmtFinSF2[is.na(test$BsmtFinSF2)] <- mean(test$BsmtFinSF2, na.rm = TRUE)
test$BsmtUnfSF[is.na(test$BsmtUnfSF)] <- mean(test$BsmtUnfSF, na.rm = TRUE)
test$TotalBsmtSF[is.na(test$TotalBsmtSF)] <- mean(test$TotalBsmtSF, na.rm = TRUE)
test$BsmtFullBath[is.na(test$BsmtFullBath)] <- mean(test$BsmtFullBath, na.rm = TRUE)
test$BsmtHalfBath[is.na(test$BsmtHalfBath)] <- mean(test$BsmtHalfBath, na.rm = TRUE)
test$GarageCars[is.na(test$GarageCars)] <- mean(test$GarageCars, na.rm = TRUE)
test$GarageArea[is.na(test$GarageArea)] <- mean(test$GarageArea, na.rm = TRUE)

# Replace missing values in categorical variables with the mode
for (i in 1:ncol(test)) {
  if (class(test[,i]) != "numeric" & sum(is.na(test[,i])) > 0) {
    test[,i] <- ifelse(is.na(test[,i]), Mode(test[,i]), test[,i])
  }
}

non_numeric_vars <- sapply(test, function(x) !is.numeric(x))
test[non_numeric_vars] <- lapply(test[non_numeric_vars], factor)
pred_test=predict(newdata=test,model0)
```

Test Data Model's Accuracy

```
tally_table_1=data.frame(actual=sub_test$SalePrice, predicted=pred_test)
mape <- mean(abs(tally_table_1$actual -tally_table_1$predicted)/tally_table_1$actual)
rmse <- sqrt(mean((tally_table_1$actual -tally_table_1$predicted)^2))
cat(paste0("RMSE: ", round(rmse, 2), "\n"))
```

```
## RMSE: 61163.74
```

```
cat(paste0("R^2: ", 0.91, "\n"))
```

```
## R^2: 0.91
```

```
cat(paste0("Accuracy: ", round(1-mape, 2), "\n"))
```

```
## Accuracy: 0.72
```

Conclusion and Remarks

In conclusion, R is a powerful tool for creating regression models and analyzing data. With a wide range of packages and functions, R can handle complex data sets and perform advanced statistical analyses. However, as with any tool, R has its limitations and drawbacks. One potential issue is the need for adequate training and experience to use it effectively. Another issue is the potential for errors in coding and data cleaning, which can lead to inaccurate results. Despite these challenges, R remains a popular choice for data scientists and analysts due to its flexibility, scalability, and open-source nature.

For this project, R was able to accurately use a training and data set to predict housing price rate changes in Kansas City to an Accuracy Percentage of 72%. Using a broader training and test set would yield a more concrete accuracy yield, but would be slower.